

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: September 26, 2006, 00:11:50 ; Search time 7.9 Seconds

(without alignments)  
2871.888 Million cell updates/sec

Title: US-10-537-002-7

Perfect score: 1560  
Sequence: 1 atgcccgcgactgcctgca.....ccaagcagcactatctgtgaa 786

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283416 segs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 200 summaries

Command line parameters:

-MODEL=frame+ n2p model -DEV=xlh  
-Q=/abs/ABSSMBR\_epool/US10537002/runat\_25092006.170505.8414/app\_query.fasta\_1  
-DB=FIR -QFMT=FASTA -SUPERX=n2p.rpr -MINMATCH=0.1 -LOOPT=0 -LOOEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human0.cdi -LIST=200  
-DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=50 -MODE=LOCAL  
-OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abss802h  
-USER=US10537002 -CGEN\_1\_1\_25 -runat\_25092006.170505.8414 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR\_80: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description                  |
|------------|-------|-------------|--------|-------|------------------------------|
| 1          | 297   | 19.0        | 280    | 2     | A39484 androgen-withdrawn    |
| 2          | 135   | 8.7         | 173    | 2     | A49182 lens membrane         |
| 3          | 133   | 8.5         | 173    | 2     | A48300 lens membrane prot    |
| 4          | 131.5 | 8.4         | 157    | 2     | G02355 tumor-associated m    |
| 5          | 126.5 | 8.1         | 160    | 2     | S21721 tumor-arrest-spec     |
| 6          | 121.5 | 7.8         | 160    | 2     | JN0503 peripher-associated m |
| 7          | 120.5 | 7.7         | 160    | 2     | JN0503 peripher-associated m |
| 8          | 118.5 | 7.6         | 1794   | 2     | T38459 hypothetrical myelin  |
| 9          | 117.5 | 7.5         | 160    | 2     | A41144 hypothetrical diver   |
| 10         | 114   | 7.3         | 695    | 2     | G64327 growth arrest-rela    |
| 11         | 107   | 6.9         | 167    | 2     | UC5044 H+-transporting tw    |
| 12         | 104.5 | 6.7         | 735    | 2     | UC5044 epithetrical membran  |
| 13         | 103.5 | 6.6         | 593    | 1     | GYHU hypothetrical prote     |
| 14         | 102.5 | 6.6         | 448    | 2     | G53610 granuln precursor     |
|            |       |             |        |       | ntpd protein - Rat           |

|    |       |     |      |   |        |
|----|-------|-----|------|---|--------|
| 15 | 102.5 | 6.6 | 466  | 2 | AD3550 |
| 16 | 102.5 | 6.6 | 1574 | 2 | T13954 |
| 17 | 101.5 | 6.5 | 530  | 2 | A45690 |
| 18 | 101   | 6.5 | 331  | 2 | C95998 |
| 19 | 101   | 6.5 | 873  | 2 | A47283 |
| 20 | 101   | 6.5 | 1964 | 2 | T09059 |
| 21 | 100.5 | 6.4 | 2318 | 2 | S45306 |
| 22 | 100.5 | 6.4 | 2318 | 2 | S45306 |
| 23 | 100   | 6.4 | 411  | 2 | PC2051 |
| 24 | 100   | 6.4 | 492  | 2 | S41288 |
| 25 | 100   | 6.4 | 1007 | 2 | T01437 |
| 26 | 99    | 6.3 | 521  | 2 | S74569 |
| 27 | 99    | 6.3 | 865  | 2 | A47282 |
| 28 | 98.5  | 6.3 | 722  | 2 | S57246 |
| 29 | 98    | 6.3 | 712  | 2 | AD2721 |
| 30 | 98    | 6.3 | 714  | 2 | H97502 |
| 31 | 98    | 6.3 | 2187 | 2 | T30826 |
| 32 | 97.5  | 6.2 | 1964 | 2 | T09059 |
| 33 | 97.5  | 6.2 | 2825 | 2 | T14271 |
| 34 | 96.5  | 6.2 | 144  | 2 | A36324 |
| 35 | 96.5  | 6.2 | 292  | 2 | A33524 |
| 36 | 96.5  | 6.2 | 775  | 1 | EDBE11 |
| 37 | 96    | 6.2 | 167  | 2 | UC5732 |
| 38 | 96    | 6.2 | 230  | 2 | A38346 |
| 39 | 96    | 6.2 | 331  | 2 | A33140 |
| 40 | 96    | 6.2 | 331  | 2 | B98147 |
| 41 | 96    | 6.2 | 361  | 2 | T45918 |
| 42 | 96    | 6.2 | 411  | 2 | PC2060 |
| 43 | 96    | 6.2 | 472  | 2 | B87658 |
| 44 | 96    | 6.2 | 595  | 2 | AF2632 |
| 45 | 96    | 6.2 | 640  | 2 | J01584 |
| 46 | 96    | 6.2 | 927  | 2 | T00357 |
| 47 | 95.5  | 6.1 | 446  | 2 | S01187 |
| 48 | 95.5  | 6.1 | 446  | 2 | T15797 |
| 49 | 95.5  | 6.1 | 691  | 2 | A25708 |
| 50 | 95.5  | 6.1 | 2531 | 2 | S45184 |
| 51 | 95.5  | 6.1 | 329  | 2 | S01075 |
| 52 | 95    | 6.1 | 341  | 2 | A13425 |
| 53 | 95    | 6.1 | 383  | 2 | D72655 |
| 54 | 95    | 6.1 | 476  | 2 | B37661 |
| 55 | 95    | 6.1 | 481  | 2 | D54463 |
| 56 | 95    | 6.1 | 2321 | 1 | KGZ0HF |
| 57 | 95    | 6.1 | 221  | 1 | KGZ0HF |
| 58 | 94.5  | 6.1 | 264  | 2 | A82766 |
| 59 | 94.5  | 6.1 | 463  | 2 | T36810 |
| 60 | 94.5  | 6.1 | 4135 | 2 | T42629 |
| 61 | 94.5  | 6.1 | 884  | 2 | A31928 |
| 62 | 94    | 6.0 | 223  | 2 | P90954 |
| 63 | 93    | 6.0 | 235  | 2 | G58803 |
| 64 | 93    | 6.0 | 337  | 2 | G95915 |
| 65 | 93    | 6.0 | 895  | 2 | S20582 |
| 66 | 93    | 6.0 | 2555 | 2 | A40043 |
| 67 | 93    | 6.0 | 572  | 2 | T29880 |
| 68 | 92.5  | 5.9 | 1520 | 2 | T00273 |
| 69 | 92.5  | 5.9 | 3020 | 2 | A43932 |
| 70 | 92.5  | 5.9 | 313  | 2 | JC5342 |
| 71 | 92    | 5.9 | 320  | 2 | A95961 |
| 72 | 92    | 5.9 | 387  | 2 | B49175 |
| 73 | 92    | 5.9 | 539  | 2 | D69748 |
| 74 | 92    | 5.9 | 718  | 2 | T02310 |
| 75 | 92    | 5.9 | 749  | 2 | T38488 |
| 76 | 92    | 5.9 | 314  | 2 | C84116 |
| 77 | 91.5  | 5.9 | 438  | 2 | T11889 |
| 78 | 91.5  | 5.9 | 524  | 2 | S88539 |
| 79 | 91.5  | 5.9 | 548  | 2 | A70780 |
| 80 | 91.5  | 5.9 | 667  | 2 | F70682 |
| 81 | 91.5  | 5.9 | 879  | 2 | B91031 |
| 82 | 91.5  | 5.9 | 1118 | 2 | A48292 |
| 83 | 91.5  | 5.9 | 1510 | 2 | T13634 |
| 84 | 91.5  | 5.9 | 3635 | 2 | T10053 |
| 85 | 91.5  | 5.9 | 349  | 2 | A10274 |
| 86 | 91    | 5.8 | 388  | 2 | AH1183 |
| 87 | 91    | 5.8 |      |   |        |

NAD(P) transhydrog  
MEGF6 protein - ra  
transactivator BEN  
probable sugar upt  
calphorin - fruit  
notch4 - mouse  
notch 3 protein -  
notch 3 protein -  
genome polypept  
hypothetrical prote  
NADH2 dehydrogenas  
calcium-binding pr  
ventral nervous sy  
H+ translocating p  
h+ translocating p  
nascent polypeptid  
notch4 - mouse  
Doc4 protein, stre  
growth arrest-spec  
high-affinity bran  
immediate-early pr  
tumor-associated m  
ultra-high-sulfur  
hypothetrical prote  
hypothetrical prote  
genome polypept  
NAD(P) transhydrog  
hypothetrical prote  
conserved hypocat  
genome polypept  
NADH2 dehydrogenas  
NADH2 dehydrogenas  
synapsin I - rat  
L-arabinose transp  
ribose transpore s  
hypothetrical prote  
probable membran  
conserved hypocat  
notch3 protein - h  
histidine/alanine-  
membrane protein X  
probable integral  
tenascin-X - bovin  
glucose transport  
hypothetrical prote  
hypothetrical prote  
probable sugar ABC  
dyctrophin-associat  
notch protein homo  
hypothetrical prote  
hypothetrical prote  
mucin 2 precursor,  
Na+/H+ antiporter  
Probable sugar upt  
Moch A protein -  
amino acid transpo  
hypothetrical prote  
trophin - human  
ribose ABC transpo  
hypothetrical prote  
disintegrin-like m  
probable membran  
probable outer mem  
mucin, tracheobron  
probable minor cai  
lamunin alpha 5 ch  
L-arabinose transp  
antibiotic resistat

|     |      |     |      |   |        |                    |
|-----|------|-----|------|---|--------|--------------------|
| 88  | 91   | 5.8 | 399  | 2 | AB0155 | probable drug resi |
| 89  | 91   | 5.8 | 400  | 2 | C97200 | uncharacterized co |
| 90  | 91   | 5.8 | 431  | 2 | AF0381 | probable ammonium  |
| 91  | 91   | 5.8 | 513  | 2 | AA4150 | structural protein |
| 92  | 91   | 5.8 | 861  | 2 | AA4825 | Notch homolog Mofc |
| 93  | 91   | 5.8 | 1700 | 2 | S08167 | Balbani ring 3 pr  |
| 94  | 91   | 5.8 | 4006 | 2 | T09070 | probable tenascin  |
| 95  | 90.5 | 5.8 | 425  | 2 | A90055 | hypothetical prote |
| 96  | 90.5 | 5.8 | 431  | 2 | B87285 | phosphate ABC tran |
| 97  | 90.5 | 5.8 | 432  | 2 | B65190 | potassium uptake p |
| 98  | 90.5 | 5.8 | 454  | 2 | B83760 | hypothetical prote |
| 99  | 90.5 | 5.8 | 455  | 2 | A11995 | amino acid transpo |
| 100 | 90.5 | 5.8 | 472  | 2 | T34748 | transmembrane tran |
| 101 | 90.5 | 5.8 | 483  | 2 | A91226 | potassium uptake p |
| 102 | 90.5 | 5.8 | 483  | 2 | G86072 | probable efpa prot |
| 103 | 90.5 | 5.8 | 530  | 2 | A70589 | MFS permease (limp |
| 104 | 90.5 | 5.8 | 542  | 2 | AF2587 | hypothetical prote |
| 105 | 90.5 | 5.8 | 542  | 2 | F97369 | mitochondrial nico |
| 106 | 90.5 | 5.8 | 1068 | 2 | T48756 | hypothetical prote |
| 107 | 90   | 5.8 | 182  | 2 | T19126 | cytochrome O ubiqu |
| 108 | 90   | 5.8 | 320  | 2 | S72756 | probable protochme |
| 109 | 90   | 5.8 | 321  | 2 | H86981 | branched-chain am  |
| 110 | 90   | 5.8 | 356  | 2 | G75444 | probable response  |
| 111 | 90   | 5.8 | 448  | 2 | G95416 | probable MFS trans |
| 112 | 90   | 5.8 | 456  | 2 | G83449 | aromatic amino aci |
| 113 | 90   | 5.8 | 457  | 1 | ORBCAA | NADH2 dehydrogena  |
| 114 | 90   | 5.8 | 736  | 2 | T13225 | histidine/alanine- |
| 115 | 89.5 | 5.7 | 218  | 2 | B25942 | hypothetical prote |
| 116 | 89.5 | 5.7 | 300  | 2 | H84194 | probable membrane  |
| 117 | 89.5 | 5.7 | 317  | 2 | F96656 | hypothetical prote |
| 118 | 89.5 | 5.7 | 472  | 2 | S28286 | cytochrome-c oxida |
| 119 | 89.5 | 5.7 | 488  | 1 | H64537 | protein C38C10.2 [ |
| 120 | 89.5 | 5.7 | 493  | 2 | G88553 | MG66 protein - ra  |
| 121 | 89.5 | 5.7 | 1574 | 2 | T13954 | notch-1 protein -  |
| 122 | 89.5 | 5.7 | 2531 | 2 | A46019 | ribose ABC transpo |
| 123 | 89   | 5.7 | 325  | 2 | D98292 | ABC transporter, m |
| 124 | 89   | 5.7 | 325  | 2 | AE2991 | multidrug resistin |
| 125 | 89   | 5.7 | 371  | 2 | D69415 | regulatory protein |
| 126 | 89   | 5.7 | 372  | 2 | T45524 | hypothetical prote |
| 127 | 89   | 5.7 | 445  | 2 | T31898 | histidine permease |
| 128 | 89   | 5.7 | 475  | 2 | E69643 | probable sulfate p |
| 129 | 89   | 5.7 | 495  | 2 | F95850 | hypothetical prote |
| 130 | 89   | 5.7 | 603  | 2 | T45093 | genome polypeptid  |
| 131 | 89   | 5.7 | 782  | 2 | S18031 | hypothetical prote |
| 132 | 89   | 5.7 | 795  | 2 | T52516 | hypothetical prote |
| 133 | 89   | 5.7 | 1737 | 2 | T00209 | MG68 protein - hu  |
| 134 | 89   | 5.7 | 1786 | 1 | MMMSB1 | laminin beta-1 cha |
| 135 | 89   | 5.7 | 3011 | 1 | GNWVCH | geneome polypeptid |
| 136 | 89   | 5.7 | 3566 | 1 | A40701 | tenascin-X precurs |
| 137 | 88.5 | 5.7 | 110  | 2 | D72701 | hypothetical prote |
| 138 | 88.5 | 5.7 | 661  | 1 | VCWVCB | cell-fate determin |
| 139 | 88.5 | 5.7 | 2471 | 2 | A49128 | hypothetical prote |
| 140 | 88   | 5.6 | 190  | 2 | A71003 | histidine/alanine- |
| 141 | 88   | 5.6 | 264  | 2 | T10468 | probable ABC trans |
| 142 | 88   | 5.6 | 333  | 2 | A10050 | methylnalonyl-CoA  |
| 143 | 88   | 5.6 | 373  | 2 | E49094 | probable transcrip |
| 144 | 88   | 5.6 | 397  | 2 | G90667 | hypothetical prote |
| 145 | 88   | 5.6 | 397  | 2 | C88518 | phenylalanine-spec |
| 146 | 88   | 5.6 | 411  | 2 | T20623 | phenylalanine-spec |
| 147 | 88   | 5.6 | 458  | 2 | A85556 | phenylalanine-spec |
| 148 | 88   | 5.6 | 458  | 2 | F90705 | phenylalanine-spec |
| 149 | 88   | 5.6 | 458  | 2 | A39431 | phenylalanine tran |
| 150 | 88   | 5.6 | 492  | 2 | G75389 | NADH2 dehydrogena  |
| 151 | 88   | 5.6 | 502  | 2 | B70845 | probable eygar tra |
| 152 | 88   | 5.6 | 543  | 2 | D83262 | hypothetical prote |
| 153 | 88   | 5.6 | 553  | 2 | F75407 | probable transpor  |
| 154 | 88   | 5.6 | 579  | 2 | T45344 | hypothetical prote |
| 155 | 88   | 5.6 | 664  | 2 | T33959 | probable urea acti |
| 156 | 88   | 5.6 | 729  | 2 | T13227 | NADH2 dehydrogena  |
| 157 | 88   | 5.6 | 752  | 2 | B82146 | Rec2-related prote |
| 158 | 88   | 5.6 | 784  | 2 | J00317 | hypothetical 82K p |
| 159 | 88   | 5.6 | 1125 | 2 | B41206 | microtubule-asso   |
| 160 | 88   | 5.6 | 2531 | 2 | T31070 | notch homolog - se |

|     |      |     |      |   |        |                     |
|-----|------|-----|------|---|--------|---------------------|
| 161 | 87.5 | 5.6 | 303  | 2 | AB3201 | conserved hypothet  |
| 162 | 87.5 | 5.6 | 306  | 2 | S08607 | chorion protein s3  |
| 163 | 87.5 | 5.6 | 330  | 2 | F64905 | probable sugar tra  |
| 164 | 87.5 | 5.6 | 330  | 2 | G85723 | probable transport  |
| 165 | 87.5 | 5.6 | 330  | 2 | B90894 | probable transport  |
| 166 | 87.5 | 5.6 | 351  | 1 | KGZQHL | histidine-rich gly  |
| 167 | 87.5 | 5.6 | 436  | 2 | A38145 | invariant surface   |
| 168 | 87.5 | 5.6 | 443  | 2 | H83589 | conserved hypothet  |
| 169 | 87.5 | 5.6 | 449  | 2 | AB2887 | MFS permease (limp  |
| 170 | 87.5 | 5.6 | 472  | 2 | F82639 | resistance protein  |
| 171 | 87.5 | 5.6 | 477  | 2 | G97662 | benzoate transport  |
| 172 | 87.5 | 5.6 | 483  | 2 | AE0914 | lirk system potassi |
| 173 | 87.5 | 5.6 | 509  | 2 | G96766 | protein lipase F2P  |
| 174 | 87.5 | 5.6 | 597  | 1 | NBHUC4 | C4b-binding protei  |
| 175 | 87.5 | 5.6 | 719  | 2 | T00266 | hypothetical prote  |
| 176 | 87.5 | 5.6 | 741  | 2 | T13791 | NADH2 dehydrogena   |
| 177 | 87.5 | 5.6 | 770  | 2 | T23999 | hypothetical prote  |
| 178 | 87.5 | 5.6 | 784  | 2 | J00317 | hypothetical 82K p  |
| 179 | 87.5 | 5.6 | 1041 | 2 | T15521 | hypothetical prote  |
| 180 | 87.5 | 5.6 | 1801 | 1 | MMRTS  | laminin beta-2 cha  |
| 181 | 87.5 | 5.6 | 3075 | 2 | S14458 | laminin alpha-1 ch  |
| 182 | 87   | 5.6 | 254  | 2 | E69853 | conserved hypothet  |
| 183 | 87   | 5.6 | 268  | 2 | S34364 | hypothetical prote  |
| 184 | 87   | 5.6 | 395  | 2 | AB1061 | probable inner me   |
| 185 | 87   | 5.6 | 413  | 2 | B83180 | probable MFS trans  |
| 186 | 87   | 5.6 | 420  | 2 | AE2856 | conserved hypothet  |
| 187 | 87   | 5.6 | 457  | 2 | D85494 | aromatic amino aci  |
| 188 | 87   | 5.6 | 457  | 2 | D90643 | aromatic amino aci  |
| 189 | 87   | 5.6 | 469  | 1 | S29126 | propeptin precuro   |
| 190 | 87   | 5.6 | 522  | 2 | S41819 | nucleoporin p62 -   |
| 191 | 87   | 5.6 | 600  | 2 | T18593 | hypothetical prote  |
| 192 | 87   | 5.6 | 2406 | 2 | A54148 | odx protein - frui  |
| 193 | 87   | 5.6 | 2515 | 1 | S47008 | tenascin-like prot  |
| 194 | 87   | 5.6 | 3011 | 2 | S40770 | genome polypeptid   |
| 195 | 86.5 | 5.5 | 166  | 2 | C72641 | hypothetical prote  |
| 196 | 86.5 | 5.5 | 346  | 2 | AH3563 | ribose transport a  |
| 197 | 86.5 | 5.5 | 397  | 2 | G69295 | oxalate/formate an  |
| 198 | 86.5 | 5.5 | 400  | 2 | AE0808 | nucleoside permea   |
| 199 | 86.5 | 5.5 | 401  | 2 | B83363 | conserved hypothet  |
| 200 | 86.5 | 5.5 | 475  | 2 | A70320 | nitrate transporte  |

## ALIGNMENTS

## RESULT 1

androgen-withdrawal apoptosis protein RVp1, prostatic - rat

C/Species: Rattus norvegicus (Norway rat)

C/Date: 03-Aug-1992 #sequence \_revision 03-Aug-1992 #text\_change 04-Mar-2000

C/Accession: A39484

R.Brüehl, M.M.; Miesfeld, R.L.

Mol. Endocrinol. 5, 1381-1388, 1991

A/Title: Isolation and characterization of transcripts induced by androgen withdrawal and

A/Reference number: A39484; MUID:92130987; PMID:1723140

C/Accession: A39484

A/Molecule type: mRNA

A/Residues: 1-280 <BRI>

A/Cross-references: UNIPARC:UPI0000163BB3; GB:W74067; NID:9205857; PIDD:AAA41760.1; PID:5

A/Genes: RVP.1

C/Superfamily: rat androgen-withdrawal apoptosis protein RVp1

## Alignment Scores:

|                        |          |               |     |
|------------------------|----------|---------------|-----|
| Pred. No.:             | 2,786-16 | Length:       | 280 |
| Score:                 | 297.00   | Matches:      | 75  |
| Percent Similarity:    | 47.8%    | Conservative: | 47  |
| Best Local Similarity: | 29.4%    | Mismatches:   | 97  |
| Query Match:           | 19.0%    | Indels:       | 36  |
| DB:                    | 2        | Gaps:         | 8   |

US-10-537-002-7 (1-786) x A39484 (1-280)

QY 28 GGGTTCGGTTCATGATGGGATGGGCGGCGCATTCGTCGACCTGATGACGACG 87

Db 9 GlythrsrleuAlaValleuGlyThrleuGlyThrleuValGlyCysCysAlaLeuPromet 28  
 Oy 88 TGAGACCCAGAGACTTGTACACACACCC--GTACAGCTGTTTCACTACACAGGCG 144  
 Db 29 TrrpValSerAlaPheIleGlySerSerIleIleThrAlaGlnIleThrTripluGly 48  
 Oy 145 CTGTGGCCCTCTGTGTCCGAGAGAGCTGTGCTTACCGAGTGGCGGGGCTTTCACC 204  
 Db 49 LeutrpMetAenCysVal--GlnSerThrGlyGlnMetGlnCysValMetCysAspSer 67  
 Oy 205 CTGTGGGGCTGTCCAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 264  
 Db 68 LeuAlaIlePheProGlnAspLeuGlnAlaAlaGlnAlaIleValValSerIleLeu 87  
 Oy 265 CTGGGTGCATTGGCTCTGTGTGTATCATCTTGTGCTGAAATGCATCCGACATTCGAC 324  
 Db 88 LeuAlaIlePheGlyLeuLeuValAlaLeuValGlyAlaGlnCysThrAsnGly--Val 106  
 Oy 325 ATGAGAGACTGTCCAG 384  
 Db 107 GlnAspGlnThrAlaValAlaValIleThrIleValAlaGlyValLeuPheLeuAla 126  
 Oy 385 GGTCTTGTGCAATGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 444  
 Db 127 AlaValleuThrleuValAlaProValSerThrSerAlaIleThrIleIleArgAspPheTyr 146  
 Oy 445 ATGTCCACAGCTAACATGTATACACCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 504  
 Db 147 --AenProLeuValProGlnAlaGlnIleValArg 156  
 Oy 505 TACACATTGTGTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 564  
 Db 157 --GlnMetGlyThrGlyLeuThrValGlyTrrAlaAlaAlaLeuGlnLeuLeuGly 175  
 Oy 565 GGTGTGATGTGTGAG 615  
 Db 176 GlnAlaLeuLeuGlyCysSerCysProProArgGlnGlyTrrAlaProThrIleValLeu 195  
 Oy 616 TACAAAGCGTTTCTTATCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 675  
 Db 196 TyrSerAla----- 198  
 Oy 676 AAGGCCAGACTGT 735  
 Db 199 ProArgSerThrGlyProGlyThrGlyThrGlyThr-----AlaTyrAspArgGlyThr 216  
 Oy 736 CGCACAG 780  
 Db 217 ThrSerGlnArgProGlyAlaArgThrProIleIleIleIleIleTyr 231  
 RESULT 2  
 A49182  
 Integral membrane protein MP20 - rat  
 N:Alternate names: lens membrane protein MP20  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 19-Dec-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C/Accession: A49182; S27883  
 R:Kumar, N.M.; Javali, L.J.; Tenbroek, E.; Louis, C.F.  
 Exp. Eye Res. 56, 35-43, 1993  
 A>Title: Cloning and expression of a major rat lens membrane protein, MP20.  
 A/Reference number: A49182; MUID:93162128; PMID:7679355  
 A/Accession: A49182  
 A:Status: Preliminary  
 A:Molecule type: nucleic acid  
 A:Residues: 1-173 <K02>  
 A/Cross-references: UNIPROT:P54825; UNIPARC:UPI000012E77B; GB:S55224; NID:g265405; PIDN:  
 A:Experimental source: lens  
 A>Note: sequence extracted from NCBI backbone (NCBIN:124760, NCBI:P124761)  
 R:Kumar, N.M.; Javali, L.J.; Tenbroek, E.; Louis, C.F.  
 Submitted to the EMBL Data Library, February 1992  
 A/Description: Cloning and expression of a major lens membrane protein, MP20.  
 A/Reference number: S27883

A/Accession: S27883  
 A:Molecule type: mRNA  
 A:Residues: 1-173 <K02>  
 A/Cross-references: UNIPARC:UPI000012E77B; EMBL:M87053; NID:g205514; PIDN:AAA41631.1  
 C:Superfamily: growth arrest-specific protein  
 C:Keywords: membrane protein  
 Alignment Scores:  
 Pred. No.: 0.00459 Length: 173  
 Score: 135.00 Matches: 48  
 Percent Similarity: 38.6% Conservative: 30  
 Best Local Similarity: 23.8% Mismatches: 78  
 Query Match: 8.7% Indels: 46  
 DB: 2 Gaps: 9  
 US-10-537-002-7 (1-786) x A49182 (1-173)  
 Oy 22 GGTGTGGGCTGT 81  
 Db 6 GlyGlyGlyLeuPhePheCysAlaIleTrrValGlyThrIleLeuLeuValAlaIleThrAlaThr 25  
 Oy 82 GACAGTGGAGACCCAG 141  
 Db 26 AspIleThrMetGlnTrrArgLeu-----SerGlySerPheAlaIleGln 40  
 Oy 142 GGGCTGTGGGCTCTGT 198  
 Db 41 GlyLeuIleTrrArgTrrCysLeuGlyAsnIleCysPheLeuGlnIleThrIleAlaTyr 60  
 Oy 199 TTCACCTGTGTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 258  
 Db 61 Trr-----AsnAlaIleThrAlaPheMetIleLeuSer 71  
 Oy 259 ATGTCTGTGGGCTGT 318  
 Db 72 AlaLeuCysAlaIleThrSerGlyIleIleMetGlyAlaLeuAla-----Phe 86  
 Oy 319 GGCAG 378  
 Db 87 AlaGlnIleSerThrPheThrArgLeuSerArgProPheSerAlaGlyIleMetPhePhe 106  
 Oy 379 GTCCTAGGCTTGT 438  
 Db 107 AlaSerThrIlePheValLeuAlaLeuAlaIle----- 118  
 Oy 439 TTCGATGTCCACAGCTAACATGTATACACCGGAGAGAGAGAGAGAGAGAGAGAG 498  
 Db 119 -----TyrThrGlyVal-----ThrValSerPheLeuGly 128  
 Oy 499 ACCAGGTAC-----ACATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 549  
 Db 129 ArgArgPheGlyAspTrrArgPheSerTrrPheTrrIleLeuGlyTrrValAlaLeuLeu 148  
 Oy 550 CTCACACTAATTTGGGGGT 594  
 Db 149 MetThrPhePheAlaGlyIlePheTrrMetCysAlaTrrArgMetHisGlnCysArgArg 168  
 Oy 595 CTGGCA 600  
 Db 169 LeuSer 170  
 RESULT 3  
 A48300  
 lens membrane protein MP19 - bovine  
 N:Alternate names: lens fiber cell membrane protein MP19; lens fiber membrane intrinsic P  
 C:Species: Bos primigenius taurus (cattle)  
 C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
 C/Accession: A48300; A34490; S05010  
 R:Gutekunst, K.A.; Rao, G.N.; Church, R.L.  
 Curr. Eye Res. 9, 955-961, 1990  
 A>Title: Molecular cloning and complete nucleotide sequence of the cDNA encoding a bovine  
 A/Reference number: A48300; MUID:91114349; PMID:2276272  
 A/Accession: A48300





```

N:Alternate names: Charcot-Marie-Tooth; GAS-3 protein; growth arrest-specific protein 3.
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 09-Jul-2004
R:Accession: JN0503; J01190; A56697; S25537
R:Edoml, P.; Martinotti, A.; Colombo, M.P.; Schneider, C.
Gene 126, 289-290, 1993
A:Title: Sequence of human GAS3/PMP22 full-length cDNA.
A:Reference number: JN0503; MUID:93246261; PMID:8482547
A:Accession: JN0503
A:Molecule type: mRNA
A:Residues: 1-160 <ED0>
A:Cross-references: UNIPROT:001453; UNIPARC:UPI0000131C8E; GB:D13203; NID:9182984; PIDN:9182984; PIDN:R.Hayasaka, K.; Himoro, M.; Nanno, K.; Sato, W.; Mura, M.; Uyemura, K.; Takahashi, E.;
R:Hayasaka, K.; Himoro, M.; Nanno, K.; Sato, W.; Mura, M.; Uyemura, K.; Takahashi, E.;
Biochem. Biophys. Res. Commun. 186, 827-831, 1992
A:Title: Isolation and sequence determination of cDNA encoding PMP-22 (PAs-II/SR13/GAS-3)
A:Reference number: J01190; MUID:93360032; PMID:1497668
A:Accession: J01190
A:Molecule type: mRNA
A:Residues: 1-160 <HAY>
A:Cross-references: UNIPARC:UPI0000131C8E; GB:D11428; NID:g220009; PIDN:BA01995.1; PID:
A:Experimental source: fetus spinal cord
R:Valentijn, L.J.; Baas, F.; Wolterman, R.A.; Hoogendijk, J.E.; van den Bosch, N.H.A.; Z
Nature Genet. 2, 288-291, 1992
A:Title: Identical point mutations of PMP-22 in Trembler-J mouse and Charcot-Marie-Tooth
A:Reference number: A56697; MUID:93265161; PMID:1303281
A:Accession: A56697
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-160 <VAL>
A:Cross-references: UNIPARC:UPI0000131C8E; GB:S61788
R:Colombo, M.
submitted to the EMBL Data Library, May 1992
A:Reference number: S25537
A:Accession: S25537
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 61-160 <COL>
A:Cross-references: UNIPARC:UPI000016A97A; EMBL:X65968; NID:g31652; PIDN:CAA46781.1; PID:
C:Genetics:
A:Gene: GDB:PMP22; GAS3
A:Cross-references: GDB:134190
A:Map position: 17p12-17p11.2
C:Superfamily: growth arrest-specific protein
C:Keywords: myelin; transmembrane protein

Alignment Scores:
Pred. No.: 0.0702 Length: 160
Score: 120.50 Matches: 46
Percent Similarity: 41.78 Conservative: 29
Best Local Similarity: 25.64 Mismatches: 56
Query Match: 7.7% Indels: 49
DB: Gaps: 8

US-10-537-002-7 (1-786) x JN0503 (1-160)
OY 61 ATCATTTGCTGCACCTGCATGACGACGACGACCCAAAGACTTGTACAAACCCCGTA 120
Db 19 LeupheValserThrlleValserGlntrp-----11e 29
OY 121 ACAGCTGTTTCAATACACGAGGCGTGGGCGTCCCTGTCGACGAGAGCTGGC--- 177
Db 30 ValGlyaNGlyHlAlaThrAspLeuTrpGlnaNCysSerThrSerSerGlyAsn 49
OY 178 -----TTACCGAGTGCCGGGCGTACTTCACCTGCTGGGGCTGCCAGCATG 225
Db 50 ValHlshIscYpPheSerSerSerProaNGlntrp----- 61
OY 226 CTGACGAGAGGCGAGCCCTGATGATCGTAGGCAACGTCCTGGGTCACATTGGCTCTG 285
Db 62 LeuGlnSerValGlnAlaThrMetIleLeuSerIleIlePheSer-----11e 77
OY 286 GTATCATCTTTCCCTGAATTCGACATTCGACATGGACAGTGAAGCACTTGCAGAGCC 345
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```

Db LeuSerLeuPheLeuPheLeuPheCys-----Gln 86

Qy 346 AACATGACATGACACCTCCGGGATCATGTCATTCTCAGGCTCTTTGGCAAT--GCT 402

Db LeuPheThrLeuThrIleGlyIleGlyArgPheTyIleThrGlyIlePheGlnIleLeuAla 106

Qy 403 GGAGGTCTGTGTTTGGCCAAACATGCTGGTGACATCACTTGGATGTGCACAGCTAACATG 462

Db GlyLeuCysVal-----MetSerAlaIleAlaIle 116

Qy 463 TACACCGGCGATGGTGGATGGTGACACTGTTCAAGCACAGTACACATTTGGTGGGCT 522

Db TyrThr--ValArgHisProGluTyrPheIleuAlaSerAspTyrSerTyrGlyPheAla 135

Qy 523 CTGTTGCGTGGGCTGGGCTGCTCGAGAGCCTTCACACTAATTTGGGGTGGATGATGATGCATC 582

Db TyrIleLeuAlaTyrValAlaPheProLeuAlaLeuLeuSerGlyValIleTyrValIle 155

RESULT 8

T38459

hypothetical divergent repeat-containing protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 09-Jul-2004

C:Accession: T38459; T38360

R:Harris, D.; McDonald, S.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, February 1996

A:Reference number: Z21794

A:Accession: T38459

A:Molecule type: DNA

A:Residues: 1-1748 <HAR>

A:Cross-references: UNIPROT:Q10172; UNIPARC:UPI0000169027; EMBL:Z69368; PIRN:CAA93290.1;

R:McLean, J.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.

submitted to the EMBL Data Library, September 1997

A:Reference number: Z21790

A:Accession: T38360

A:Molecule type: DNA

A:Residues: 1457-1794 <MC1>

A:Cross-references: UNIPARC:UPI000016208D; EMBL:Z70691; PIRN:CAA94638.1; GSPDB:GN00066; S

C:Genetics:

A:Experimental source: clone C25G10

A:Gene: SPAC27F1.01C; SPDB:SPAC25G10.09C

A:Map position: 1

Alignment Scores:

| Pred. No.:             | Length: | Matches: | Conservative: | Mismatches: | Indels: | Gaps: |
|------------------------|---------|----------|---------------|-------------|---------|-------|
| 0.0837                 | 1794    | 73       | 37            | 99          | 83      | 15    |
| Score:                 | 118.50  | 7.6%     | 25.1%         | 7.6%        |         |       |
| Percent Similarity:    |         |          |               |             |         |       |
| Best Local Similarity: |         |          |               |             |         |       |
| Query Match:           |         |          |               |             |         |       |

DB: 2

US-10-537-002-7 (1-786) x T38459 (1-1794)

Qy 749 TCGTCTCTGTCGCGGACACCTCATGTAATCTTGTGTTTGGTTGGACCCAAAG 690

Db 1372 SerSerAsnValHisIleProProProProValGlnProMetAlaMetProSerHis 1391

Qy 689 CCAGTGCTGACCTTGAAGCCTCCAGGCTTGTAGCAACACTGTGGCCTGAGCGATGATA 630

Db 1392 AsnAlaValAsnAlaArgProSerAla-----ProGluArg----- 1403

Qy 629 GAAAGCGCTTGTAGTGTGTTCTTCTGTGACCAGGCCCGGACAGGAGATGCATCATC 570

Db 1404 ArgAspSerPheGlySerValSerSerGlySerAsn-----ValSerSerIle 1419

Qy 569 ACACCCCAATTAGTGTGAGGCT--CCAGCAGCCAGCCCAAGAACAGAGCCGACCA 513

Db 1420 GluArgGluThrSerThrMetProLeuLysAlaSerGlnProThrAsnProGlyAlaPro 1439

Qy 512 AATGTGTACTTGCTGTGAACAGTGTGCACCATCCACCATGCGCGGTATCATTTAGCT 453

Db 1440 SerAsnHis-----AlaProGluValAlaProProAlaProMetHisAlaValAla 1456



## Alignment Scores:

|                        |        |               |     |
|------------------------|--------|---------------|-----|
| Pred. No.:             | 0.211  | Length:       | 695 |
| Score:                 | 114.00 | Matches:      | 54  |
| Percent Similarity:    | 42.7%  | Conservative: | 39  |
| Best Local Similarity: | 24.8%  | Mismatches:   | 75  |
| Query Match:           | 7.3%   | Indels:       | 50  |
| DB:                    | 2      | Gaps:         | 11  |

US-10-537-002-7 (1-786) x G64327 (1-695)

```

25 TTGGGGTTTCGGGTTTCACTGATTTGGGATTTGCG-----GGCATCATTTGCTGCG 72
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db LeuGIlyTylLeuThrLeuThrlaGlylleSerThrValIleMeGlylleIleThrGly 444
QY 73 ACCTGCATGAGCAGCAGTGCAGACCCCAAGACTTGTACACACCCCGTAA-----123
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 445 GlyTylLeuGlyAspPhe---ThrTylGlyPhePheGlyPheAspValThrIlyThrPro 463
QY 124 ---GCTGTTTCACTGACAGGAGGCTGTGGCGCTCGCTGTCGAGAGAGCTGTGCTTC 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 464 LeuAlaLeuValAspPheLeuGly-----GlySerTylTylIle 476
QY 181 ACCGATGCGCGGCGTACTTCACTGCTGCGGCTGCGAGCCATGCTGCGAGCAGTGCAG 240
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 477 AsnAsnAsnAsnProLeuPheThr---LeuGlySerIleSerValThrAsnGlyPromet 495
QY 241 GCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 496 AlaIleLeuValPheSerIlePheValGlyLeuIleIleIleIleIleIleIleIlePhe--- 514
QY 301 CTGAATGATCGGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCG 360
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 515 ---ValGlyPheIlyGlyAsnValIlyAspGlyAspMet-----526
QY 361 TCCGGATCATGTTCACTTGTCTGAGTCTTTGTCGATTTGTCGATTTGTCGATTTG 417
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 527 ---GlyAspAlaPheIleAsnGlyValTyrIleLeuLeuIleLeuSerIlePheVal 545
QY 418 -----GCCAACATGCTGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 450
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 546 GlyIleGlyLeuMetPheAlaGlyAlaAsnThrMetIleAlaGlyIleIleIleIle 565
QY 451 -----ACAGCTAACATGATACACCGGC-----ATGGTGGAGATGATG 486
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 566 PheValIleValLeuAlaIleLeuAlaSerMetTyrIlyGlyTyrIlySerGlyValIleMet 585
QY 487 CAGACTGTTACAGACAGGATACACTTTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTG 546
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 586 GluAlaIle-----LeuGlyAlaMetAspValThrGlyPheLeuGlyAsn 600
QY 547 GGCCTTCACACTAATTTGGGCTGTGATGATGATGATGATGATGATGATGATGATGATG 600
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 601 ValLeuSerTyrAlaArgLeuLeuAlaLeuCySLeuAlaThrGlyIleValIleAla 618

```

## RESULT 11

JC5044

epithelial membrane protein 2 - human

C:Species: Homo sapiens (man)

C:Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 09-Jul-2004

C:Accession: JC5044

R:Taylor, V.; Suter, U.

Gene 175, 115-120, 1996

A:Title: Epithelial membrane protein-2 and epithelial membrane protein-3: Two novel mem

A:Reference number: JC5044; MUID:97074659; PMID:8917086

A:Accession: JC5044

A:Molecule type: mRNA

A:Residues: 1-167 &lt;TAY&gt;

A:Cross-references: UNIPROT:P54851; UNIPARC:UPI000016AP4; EMBL:X94770; NID:91359880; PI

C:Comment: This protein belongs to peripheral myelin protein 22 family which are involve

C:Superfamily: growth arrest-specific protein

Alignment Scores:

|                        |        |               |     |
|------------------------|--------|---------------|-----|
| Pred. No.:             | 0.881  | Length:       | 167 |
| Score:                 | 107.00 | Matches:      | 47  |
| Percent Similarity:    | 39.0%  | Conservative: | 29  |
| Best Local Similarity: | 24.1%  | Mismatches:   | 73  |
| Query Match:           | 6.9%   | Indels:       | 46  |
| DB:                    | 2      | Gaps:         | 10  |

US-10-537-002-7 (1-786) x JC5044 (1-167)

```

25 TTGGGGTTTCGGGTTTCA-----CTGATGGGATTTGGGCGATTTGCTGCCACTGCG 78
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 LeuAlaPheIleIleAlaPheHisIleThrSerAlaIleLeuLeuIleAlaThrVal 24
QY 79 ATGACGACGTGAGACCCCAAGACTTGTACACACCCCGTAAAGCTGTTTCAACTAC 138
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 25 AspAsnAlaTyrIleValGlyAspIlePhe-----Phe 35
QY 139 CAGGAGCTGTGGCGCTCTGTGTGTCGAGAGAGCTGTGCTGCTGCTGCTGCTGCTGCTG 189
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 36 AlaAspValTyrPheGlyCysThrAsnAsn-----ThrAsnCyThrValIle 51
QY 190 -----CGGGGCTACTTCACTGCTGCGGCTGCGAGCCATGCTGAGGCAAGT 237
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 52 AsnAspSerPheGlnGlyIlyrSerThr-----LeuGlnAlaPhe 64
QY 238 CGAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 297
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 65 GlnAlaThrMetIleLeuSerThrIleLeuGlyCysIleAlaPhePhe-----IlePhe 82
QY 298 GCCCTGAATGATCGGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATGCGATG 357
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 83 ValLeuGlnLeuPheArgLeuGlyGlnGlyIle-----ArgPheValIleu 97
QY 358 ACCTCCGGATCATGTTCACTTGTCTGAGTCTTTGTCGATTTGTCGATTTGTCGATTTG 417
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 98 ThrSer---IleIleGlnLeuMetSerCysLeuGlyValMetIleAlaIleSerIleTyr 116
QY 418 GCCAACATGCTGCTGATCTGATGATCTGATGATCTGATGATCTGATGATCTGATGATCTG 477
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 117 ThrAspAlaGlyG---GlyAspIleHisAspIlyAsnAlaIlyrPheTyr-----131
QY 478 GGGATGTGACAGACTTTCAGACAGGATACACTTTGCTGCGGCTGCTGCTGCTGCTGCTG 537
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 132 -----ProValThrArgGlnGlySerTyrGlyTyrSerTyrIleLeuAlaTyr 147
QY 538 CTGCGTGAAGGCTCACAATAATTTGGGCTGTGATGATGATGATGATGATGATGATGATG 592
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 148 ValAlaPheAlaCySThrPheIleSerGlyMetCysTyrLeuIle 162

```

## RESULT 12

T45059

hypothetical protein Y39B6B.gg [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004

C:Accession: T45059

R:Wilson, R.; Alnecough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burton, C.

raser, A.; Pulton, L.; Gardner, A.; Green, P.; Hawkins, T.; Hillier, L.; Jier, M.; Johnst

B.; O'Callaghan, M.; Parsons, J.; Percy, C.; Riken, L.; Roopra, A.; Saunders, D.

Nature 368, 32-38, 1994

A:Authors: Showkhen, R.; Sims, M.; Smaldon, N.; Smith, A.; Smith, M.; Sonhammer, E.; St

tock, L.; Wilkinson-Spyrat, J.; Wohldman, P.

A:Title: 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.

A:Reference number: S43531; MUID:94150718; PMID:7906398

A:Accession: T45059

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-735 &lt;WIL&gt;

A:Cross-references: UNIPROT:O9NES7; UNIPARC:UPI0000798CB; EMBL:AL132896; NID:96434440; I

C:Experimental source: clone Y39B6B

A:Map position: 3

A:introns: 18/1, 69/1

A:Note: Y39B6B.gg





Query Match: 6.6% Indels: 194  
 DB: 1 Gaps: 16  
 US-10-537-002-7 (1-786) x GYHU (1-593)

```

  QY 57 GGGCATCATTTGCTGCACCTGCATGAGCAGTGGAGCAGCCAGACTTGTACAGAACCC 116
  DB 95 G|YH|S|H|S|C|Y|S|C|S|-----Pro 100
  QY 117 CGT-----AACAGCTGTTTCACTACAGAGG-----143
  DB 101 ArgGlyPheH|S|C|Y|S|e|S|e|L|A|a|s|p|G|Y|A|g|S|e|S|e|P|H|e|G|I|A|g|S|e|G|Y|A|s|n|a|n|S|e|r 120
  QY 144 GCTGTCGGCCTCTGTCGTCGAGAGCTCTGCTTCAACGAGCGCGGCTACTCTTCAAC 203
  DB 121 ValG|Y|A|I|I|e|G|I|n|C|Y|S|P|P|o|A|s|P|S|e|I|n|-----P|H|e|G|I|C|Y|S|P|P|o|A|s|P|P|H|e|S|e|T 138
  QY 204 CCTGC-----TGGGCTGCAGCCATGCTGCAGGAGCTGCAGC 242
  DB 138 hrC|Y|S|C|Y|e|V|a|I|e|t|V|a|s|p|G|Y|S|e|T|P|G|Y|C|Y|S|P|P|o|M|e|T|P|G|I|n|A|A|S|e|S|e|C|Y|S 157
  QY 243 CCTGATGATCCTAGGCAT-----CGTCCCTGGG 269
  DB 158 C|Y|S|G|I|A|S|P|A|T|G|V|A|H|I|S|C|Y|S|P|P|o|H|I|S|G|Y|A|A|P|H|e|C|Y|S|A|S|P|L|e|U|V|A|H|I|S|T|H|A|I|g 177
  QY 270 TGCCATTTGGCCTCTGTCATTCAT-----293
  DB 178 C|Y|S|I|e|T|H|P|P|o|T|H|G|Y|T|H|H|I|S|P|P|o|L|A|I|V|S|Y|S|L|e|U|P|P|o|A|I|A|H|g|H|H|S|H|e|n 197
  QY 294 -----CTTTCCTGGAATGATCCGATTTGGAGCAT 326
  DB 198 ArgAlaValAlaLeuSerSerSerValMetC|Y|S|P|P|o|A|S|P|A|A|A|g|S|e|A|g|S|P|P|o|A|S|P 217
  QY 327 GGA-----GACTCTGCCAAGCCAACTAGAC 353
  DB 218 G|Y|S|e|S|e|T|H|P|P|o|S|e|L|Y|S|Y|G|Y|C|Y|S|P|P|o|M|e|P|P|o|S|n|A| 237
  QY 354 ACTGACCTCCGGATCATTTCTTGT-----CTCAGG 386
  DB 238 T|H|C|Y|S|S|e|S|e|A|P|H|I|e|U|H|I|S|C|Y|S|P|P|o|I|n|A|S|P|H|I|V|A|C|Y|A|S|P|L|e|U|I|e|G|I|n 257
  QY 387 TCTTTGTC-----395
  DB 258 S|e|L|Y|S|C|Y|S|L|e|U|S|e|L|Y|G|U|S|n|A|A|T|H|T|H|A|S|P|L|e|U|T|H|Y|S|L|e|U|P|P|o|A|H|I|S 277
  QY 395 -----395
  DB 278 T|H|V|A|I|G|Y|A|S|P|V|A|L|Y|S|A|S|P|M|e|T|G|U|V|A|S|e|r|C|Y|S|P|P|o|A|S|P|G|Y|T|Y|T|H|C|Y|S|C|Y|S 297
  QY 396 -----ATTGCTGAGATGTCGT-----GTTTGCACATGCT 428
  DB 298 ArgLeuG|I|n|S|e|L|Y|A|I|T|P|G|Y|C|Y|S|P|P|o|H|e|T|H|I|n|A|I|V|A|C|Y|S|-----314
  QY 429 GGTGACTAATTTGATGTCACAGCTAACATGTACACCGGATGGGTGGATGTGCA 488
  DB 314 -----CTTTCCTGGAATGATCCGATTTGGAGCAT 514
  QY 489 GACTGTTCAAGCAGTACAC-----ATTGGTGGGCTCTGTTGTCGG 533
  DB 315 ---C|Y|S|G|I|A|S|P|H|I|S|I|e|H|I|S|C|Y|S|P|P|o|A|G|Y|P|H|e|T|H|C|Y|S|A|S|P|H|I|n|G|Y|S|G|Y 333
  QY 534 CTGGGTCCTGGAGCCTCAC---ACTAATTTG-----563
  DB 334 T|H|C|Y|S|G|I|U|G|I|n|G|Y|P|P|o|H|I|S|G|I|n|V|A|P|P|o|T|P|P|M|e|T|G|U|V|A|S|P|P|o|A|H|I|S|L|e|U|S|e|r 353
  QY 564 -----GGGTGATGATGTCATGTCCTGTCGG 590
  DB 354 L|e|U|P|P|o|A|S|P|P|o|G|I|n|A|I|e|U|Y|A|S|P|V|A|P|P|o|C|Y|S|A|S|P|A|n|V|A|S|e|S|e|r|C|Y|S|P|P|o 373
  QY 591 GGGCTGGACCAAGAAACCACTACAAAGCCTTTCTTATCATGCTCAGGCAACAG 650
  DB 374 S|e|r|S|e|A|S|P|H|I|n|C|Y|S|G|I|n|L|e|U|T|H|S|e|r|G|Y|U|T|P|G|Y|C|Y|S|A|S|P|P|o|I|e|P|P|o|G|I|n 393

```

QY 651 -----TGTGCTTCAAGCCTGG 668  
 DB 394 AlaValC|Y|S|S|e|S|e|A|P|H|I|S|I|e|H|I|S|C|Y|S|P|P|o|I|n|G|Y|T|Y 408

RESULT 14  
 G53610  
 ntp protein - Enterococcus hirae  
 C|Species: Enterococcus hirae  
 C|Date: 07-Jul-1995 #sequence\_revision 07-Jul-1995 #text\_change 18-Sep-1998  
 C|Accession: G53610  
 R|Takase, K.; Kakimura, S.; Yamato, I.; Konishi, K.; Igatahshi, K.; Kakimura, Y.  
 J. Biol. Chem. 269, 11037-11044, 1994  
 A|Title: Sequencing and characterization of the ntp gene cluster for vacuolar-type Na(+)-  
 A|Reference number: A53610; MUID:94209269; PMID:8157629  
 A|Accession: G53610  
 A|Status: preliminary  
 A|Molecule type: DNA  
 A|Residues: 1-448 <TAK>  
 A|Cross-references: UNIPARC:UPI0000178EC9; GB:D17462  
 C|Superfamily: Na+-ATP synthase chain J

Alignment Scores:  
 Pred. No.: 1.89 Length: 448  
 Score: 102.50 Matches: 47  
 Percent Similarity: 44.1% Conservative: 28  
 Best Local Similarity: 27.6% Mismatches: 64  
 Query Match: 6.6% Indels: 31  
 DB: 2 Gaps: 10

US-10-537-002-7 (1-786) x G53610 (1-448)

```

  QY 139 CAGGGGCTGTGGCGCTGTCGTCGAGAGAGCTCTGCTTCAACGAGTGCAGGCGGCTTAC 198
  DB 157 L|Y|S|G|Y|I|e|T|P|P|H|e|S|e|I|P|H|e|H|I|S|A|V|A|S|e|r|P|H|e|-----C|Y|A|S|n|A|I|e|I|Y 174
  QY 199 TTCACCTGTCGGG-----CTGCCACCATGCTGCAGGAGTGCAGGCGGCTGATGATC 252
  DB 175 P|H|e|A|S|P|L|e|U|G|Y|A|S|P|S|e|r|L|e|U|H|A|S|P|G|I|n|T|H|A|n|V|A|I|Y|L|e|U|I|e|W|e|T|V|A| 194
  QY 253 GTAAGC-----ATGTCCTGGTGCCTTGGCCCTCCGTCGATTCATTTGCCCTGA 306
  DB 195 V|A|S|e|r|A|L|e|U|I|e|I|e|A|I|e|A|G|Y|L|e|U|I|P|H|e|I|V|A|I|T|P|A|S|P|I|L|e|U|S|e|r 214
  QY 307 TGATCCGATTTGGAGCATGAGAGTCTGTCGCAAGCCAAAGCCAACTAGACCTGC--- 363
  DB 215 T|Y|H|S|A|S|P|V|A|I|Y|S|I|e|T|H|L|e|U|H|I|S|e|r|Y|S|V|A|L|A|L|e|U|S|e|r|V|A|I|T|H|A|L|e|U 234
  QY 364 -----GGATCATGTTCTTGTCTC-----TCAAGTCTTTGTGCAAT 399
  DB 235 L|e|U|e|U|I|e|G|Y|G|Y|P|H|e|I|L|e|U|P|H|e|U|I|e|T|H|G|U|A|T|G|A|n|G|Y|L|e|U|T|H|L|e|U|V|A| 254
  QY 400 GCTGAGTGTCTGTCGTCGTCGCAACATGTCGTGACTACTTCTGATGTC-----450
  DB 255 L|Y|S|G|Y|-----T|H|P|H|e|T|H|G|U|A|T|G|L|e|U|A|A|A|n|T|H|P|H|e|M|e|r|S|e|r|V|A|I|T|H|P|P|o 272
  QY 451 ---ACAGCTAACATGTACAC-----GGCATGGGAGGATGGTCAG 489
  DB 273 A|R|G|H|I|A|G|Y|T|Y|T|S|e|r|I|L|e|A|S|P|Y|I|L|e|U|I|n|M|e|T|S|e|r|H|I|S|A|I|e|G|Y|L|e|U|I|e|U 292
  QY 490 ACTGTTCAAGCAGTACACATTTGGTGGGCTCTGTTGTCGTCGTCGTCGTCGAGGC 549
  DB 293 T|H|W|e|T|P|H|e|U|M|e|T|Y|I|L|e|G|Y|G|Y|T|H|S|e|r|-----G|Y|S|e|r|T|H|A|I|G|Y|G|Y 309
  QY 550 CTC---ACACTAATTTGGGCTGTGATGATG 576
  DB 310 L|e|U|Y|S|T|H|T|H|T|H|L|e|U|G|Y|I|L|e|U|e|U 319

```

RESULT 15  
 AD3550  
 NMD(P) trianhydrogenase (AB-specific) (EC 1.6.1.2) - Brucella melitensis (strain 16M)  
 C|Species: Brucella melitensis  
 C|Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 31-Dec-2004  
 C|Accession: AD3550

R:delVecchio, V.G.; Kaparaju, V.; Redkar, R.J.; Patra, G.; Mujet, C.; Los, T.; Ivanova, M.; Mazur, M.; Coleman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Leese, Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A>Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
 A/Reference number: AD3252; PMID:11756688  
 A/Accession: AD3550  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-466 <KUN>  
 A/Cross-references: UNIPROT:Q8YD52; UNIPROT:Q8FV76; UNIPARC:UPI000058434; GB:AE008918;  
 A/Experimental source: strain 16M  
 C/Genetics:  
 A/Gene: BME10325  
 A/Map position: 11  
 C/Superfamily: NAD(P) transhydrogenase, beta subunit; NAD(P) + transhydrogenase (B-specific C/Key words: oxidoreductase

Alignment Scores:  
 Pred. No.: 1.88 Length: 466  
 Score: 102.50 Matches: 65  
 Percent Similarity: 34.6% Conservative: 38  
 Best Local Similarity: 21.8% Mismatches: 76  
 Query Match: 6.6% Indels: 119  
 DB: 2 Gaps: 15

US-10-537-002-7 (1-786) x AD3550 (1-466)

```

OY 28 GGGTTCGTGTTTCACTGATGGATTGCG-----GGCATC----- 63
DB 58 GYLeuValLeuIleIleValGlyIleAlaIleGlyGlyIleGlyValAlaIleAla 77
OY 64 -----ATTGCTGCCAAGCTGACGACGAGACCCCAAGACTGTGACCAACCC 117
DB 78 ArgArgIleAlaMetThrAlaMetProGln----- 87
OY 118 GAAACAGCGTTTTCACACACAGGGGGCTGGCGCTCTGTCCGAGAGACTGTGC 177
DB 88 LeuValAlaAlaPheHis----- 93
OY 178 TTCACGAGTGGCGGGCTACTTCACTGCTGGGGCTGCCAGCATGTCGACGAGTG 237
DB 94 -----SerLeuIleGlyLeuAlaAlaValLeuValAlaAla 105
OY 238 CGAGCCCTG----- 246
DB 106 AlaAlaLeuTySerProHisSerPheGlyIleGlyValGlyAlaIleHisGlyGln 125
OY 247 -----ATGATCGTAGCATCGCTGGGGGCGCATGGGCTCTGATTCATC 294
DB 126 AlaLeuIleGlyMetSerLeuGlyValAlaIleGlyAlaIleThrPheThrGlySerIle 145
OY 295 TTTCGCTGGAATGATCGCATC-----GGCAGCATGAGAGACTGCCAAGCCAATG 351
DB 146 ILeAlaPhe-----LeuIleLeuAspGlyArgMet----- 155
OY 352 ACACTGACCTCCGGGATCATGTTCACTGTCTCAGCTCTT-----TGTGCATT 399
DB 156 -----SerGlyLeuProIleMetLeuProGlyArgHisValIleAsnAlaIleLeu 172
OY 400 GCTGAGTGTCTGTGTTGGCAACATGCTGGAGTAAC-----TTT 441
DB 173 AlaAlaAlaIleValLeuLeuIleValIleThrIleThrIleSerHisPheValPhe 192
OY 442 TGGATGTCACAC-----GCTAATCATGACCGCGCATGGGCGG 480
DB 193 TrpLeuIleValLeuLeuSerLeuValPheGlyValLeuIleIleValProIleGlyGly 212
OY 481 -----ATGTGAGACTGTTCAGACAGTACATTTGGCGGCTCTGTTT 528
DB 213 AlaAspMetProValValValSerMetLeuAsnSerTySerGlyTrpAlaAlaIleGly 232
OY 529 GTGGCTGGGTGGCGAGGCTCACATTAAT----- 561

```

```

DB 233 ILeGlyPheThrLeuGlyAsnLeuAlaLeuIleIleThrGlyAlaLeuValGlySerSer 252
OY 562 GGGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 621
DB 253 GlyAlaIleLeuSerTyIleMetCysIleGlyMet-----AsnArgSer 267
OY 622 GCGGTTCTTATCATGCTCCAGGACAGTGTGCTCAAGCTGAGGCTTCAAGGCC 681
DB 268 PheIleSerValIleLeuGlyIlePhe-----GlyGlyAspThr 280
OY 682 AGCATGCTGTTGGTTCACACCAAAACAGAGATATACATGATGAGTGGCC 735
DB 281 SerSerAlaAlaGlyGlyValGlyValGlyValGlyValGlyValGlyValGlyVal 298

```

RESULT 16

MEGF6 protein - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
 C/Accession: T13954  
 R/Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.  
 Genomics 51, 27-34, 1998  
 A/Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs  
 A/Reference number: Z14126; MUID:98360089; PMID:9693030  
 A/Accession: T13954  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-1574 <NAK>  
 A/Cross-references: UNIPROT:O88281; UNIPARC:UPI000043BEE; EMBL:AB011532; NID:93449293;  
 A/Experimental source: strain Sprague-Dawley; brain  
 C/Genetics:  
 A/Gene: MEGF6

Alignment Scores:

Pred. No.: 1.7 Length: 1574  
 Score: 102.50 Matches: 60  
 Percent Similarity: 27.0% Conservative: 14  
 Best Local Similarity: 21.9% Mismatches: 83  
 Query Match: 6.6% Indels: 117  
 DB: 2 Gaps: 16

US-10-537-002-7 (1-786) x T13954 (1-1574)

```

OY 36 GGTTCACATGATGGATTTGGGCGCATGTTGTCACATGACATGACATGACATGAC 95
DB 725 GlyTyGlnGlyGlyAspCysGlyGlnGlyCys-----ProValGlyThr 739
OY 96 CCAAGACTTGTACAAACACCCGTAACAGCTGTTTCACTACACAGGGGCTGTGGCGCTC 155
DB 740 PheGlyValAlaGlySerGlySerCysValGlyAlaPro----- 754
OY 156 CTGTGTCGAGAGAGACTGTGCTTACCGAGTCCGGG-----CTACTTACCTGTGG 212
DB 755 -----CysHisArgValThrGlyGlyCysLeuGlySerProGly 767
OY 213 GCTGCCACCATGCTGACGAGCGAGCCCTGATATGTAAGGACATGCTCTGGGTC 272
DB 768 LysThrGlyGlyAspCysGlyAlaAspCysProGlyGlyArg----- 781
OY 273 CATTCGCTCTGATGATCATCTTTCGCTGAAATGATCCGATGACATGACATGACATG 332
DB 782 ---TrpGlyLeuGly-----CysGlnGlnIleCysProAlaCysGlnHisGlyAla 797
OY 333 CTTCGCAAGCAACATGACATGACATGACATGACATGACATGACATGACATGACAT 383
DB 798 SerCysAsnProGlnThrGlyThrCysLeu-----CysLeuProGlyPhe 812
OY 384 -----ArgCysGlnThrGlyThrCysLeu-----TGTGAGT 407
DB 813 ValGlySerArgCysGlnAspThrCysSerAlaGlyTyrGlyThrGlyCysGlnIle 832
OY 408 GTCTGTGTTTGCACATGCTGTGATCTTGTGATGTCACAGTAAATGATGATAC 467

```



Score: 101.00  
Percent Similarity: 29.3%  
Best Local Similarity: 23.6%  
Query Match: 2  
Matches: 45  
Conservative: 30  
Mismatch: 64  
Indels: 52  
Gaps: 8

US-10-537-002-7 (1-786) x C95998 (1-331)

QY 28 GGGTCGTCGTCCTCACTGATTGGGATTCGGGACATC-----GCTCCACCTGC 78  
DB 126 GYPLEPHEVALTHRARGILEGLYVALAENALAEULIETHRTHLEUGLYTHREULA 145  
QY 79 ATGACACAGTGAAGACCCAGACTGTGACAAACCCCGTAACAGCTGTTTCAACTG 138  
DB 146 IIEPHEAGLYLEUTHRGLINSERIIEGLYGLYARGANILEPROILEALANPHE 165  
QY 139 CAG---GGGCTGCGCTCTCTGTGTCGAGAGAGCTGTGCTTACCGAGTCCGGGCG 195  
DB 166 IEPTRPALLIETPRARGPRO----- 172  
QY 196 TACTTCACCTCTCTGCGGCTGCGACGCTGCGAGGAGCTGATGATCTGA 255  
DB 173 PHELEUANILEPROLEUSERALEUVALPHELEUVALALAELET---VAL 190  
QY 256 GGCATGTCCTGTCGTCCTGCTCCGATCCATCTTGCCTGAATGATCCGC 315  
DB 191 GYILEVALLEULYARG-----SERVALPHEGLYARGSERILETYRALA 205  
QY 316 ATGGCAGACGATGAGACTCTGCAAGCCAAAGCAATGACACTGCTCCGGATCATGTTG 375  
DB 206 IIEGLYALAEUVALAENALAEULIETHRTHLEUGLYLEUGLYLEULIETHRTHLE 223  
QY 376 ATGTGTCAGGCTCTTGTGCAATTGCTGAGTGTCTGTGTTCCAACTGCTGTACT 435  
DB 224 -----VALPHEALIGLYPHELEU----- 230  
QY 436 AACTCTGGATGTCACAGACTAATGATACCGGAGTGGTGGATGTCAGACTGT 495  
DB 231 -----SERGLYALPHEIIEGLYLEUGLYLEULIETHRTHLE 244  
QY 496 CAGACGAGTACACTTGTGCG-----GCTCTGTTCTGTGCG 534  
DB 245 GINLEUGLYSERTHSERGLYTHRTHGLYLEUGLYLEUGLYLEULIETHRTHLE 264  
QY 535 TGGGTGCTGAGGCTCTCACTAATTTGGGGGT 567  
DB 265 VALILEUGLYLYTHRTHSERLEUGLYLY 275

RESULT 19  
A47283  
calphoctin - fruit fly (Drosophila melanogaster)  
CISpecies: Drosophila melanogaster  
CDate: 21-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
CAccession: A47283  
R/Ballinger, D.G.; Xue, N.; Harehman, K.D.  
P/Boc Natl. Acad. Sci. U.S.A. 90, 1536-1540, 1993  
A/Title: A Drosophila photoreceptor cell-specific protein, calphoctin, binds calcium and  
A/Reference number: A47283; MUID:93165730; PMID:8434015  
A/Contents: photoreceptor cells  
A/Accession: A47283  
A/Status: preliminary  
A/Molecule type: nucleic acid  
A/Residues: 1-873 <BAU>  
A/Cross-references: UNIPROT:Q02910; UNIPARC:UPI000016B846; GB:I05080; NID:G157071; PIDN:  
C/Note: sequence extracted from NCBI backbone (NCBIN:124958, NCBI:124959)  
C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;

Alignment Scores:  
Pred. No.: 2.37 Length: 873  
Score: 101.00 Matches: 60  
Percent Similarity: 36.2% Conservative: 34  
Best Local Similarity: 23.1% Mismatches: 100  
Query Match: 6.5% Indels: 66

DB: 2 Gaps: 12  
US-10-537-002-7 (1-786) x A47283 (1-873)

QY 695 CCMAAGCAGTGTGCTTGAAGCT---CCAGCTGTAGCAACCTGTGCTGAG 639  
DB 370 PROIALPROVALIALATHRTHPROVALPROIALTHREULIETHRTHASPROALP 389  
QY 638 GCATGATGAAGAGGCTTTGATGTTGTTCTTCTGTGCTCCAGGCCCCGAGCGATG 579  
DB 390 -----VALTHRALSERIALVALPRO---GLUEUPRO 399  
QY 578 CACATCATCACACC-----CAATTAGTGAAG 549  
DB 400 PROVALIIEALPPOSERPROVALPROSERIALVALIAGLUTHPROVALIIEPHEULA 419  
QY 548 CCTCCAGGAGCCAGCC---ACGAACAGAGCCGACCAATGTTACCTGTGTAACA 492  
DB 420 PROPROVALIIEUPROPROVALIIEALAGLUPROVALIIEALAGLUTHR 439  
QY 491 GTCGACACATCCACCC-----ATGCCGTGTACATGTTAGCTGTGACATCCAGAAG 438  
DB 440 PROGLUTHRPROIALPROIALSERIALPROVALIIEALAEULIETHRTHLEULIETHRTH 459  
QY 437 TTAGTCACAGACATGTGGCAACACAGACACTCCAGCA----- 399  
DB 460 VALIAPROVALIIEALALAPROSERIALPVALIIEALAGLUALAPROSERIALALALA 479  
QY 398 ---ATTGCAACAAAGCTGAGACATGATCCGAGTGGTGTGCT 342  
DB 480 PROIIEALSERTHRTHPROIIEALSERVALPROGLUTHRTHALAPROIALA 499  
QY 341 TTGGCAGAGTCTCTCACTGCAATGCGATGATTTCAAGGCAAAATGATTCAGG 282  
DB 500 ALAVALPROTHGLUPROIALASPVALSERVALIIEULIETHRTH 518  
QY 281 AGCCCATGAGCAGCCAGAGACGATGCTGATCATGACAGGCTGCTGCTGACATG 222  
DB 519 ---PROVALIIEALP-----PROVALIIEULIETHRTHGLIETHRTHGLIETHRTH 534  
QY 221 GCTGACAGCCCGACAGAGGTGAG-----TAGCCCGGACACTG 183  
DB 535 VALIAPROPROGLUALIIEALAGLUPHEULIETHRTHGLIETHRTHGLIETHRTH 554  
QY 182 GTGAAGCAGAGCTTCTCGACACAGAGAGCCAGCCCTGTGTTGAAACAGT 123  
DB 555 IIEPROALPHEULIETHRTH-----THSER 564  
QY 122 GTTACGGGAGTGTGATACAGCTTGGGTCCACTGTGTCATGACAGTGGACAGT 63  
DB 565 VALPROIALA-----GLIETHRTHGLIETHRTH 572  
QY 62 ATGCCCGATCCCATGATGTAACCAAGACCCCAAGCCCTGACAGGACTGAGCC 3  
DB 573 SERTHRTHSERPROIALPROGLIETHRTHSERLEUPROPROALAGLUALIETHRTH 592

RESULT 20  
T09059  
notch4 - mouse  
CISpecies: Mus musculus (house mouse)  
CDate: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
CAccession: T09059  
R/Rosen, L.; Mahatras, G.; Qin, S.; Ahern, M.E.; Dankers, C.; Laeky, S.; Loretz, C.; Sci  
submitted to the EMBL Data Library, October 1997  
A/Description: Sequence of the mouse major histocompatibility locus class III region.  
A/Reference number: Z16543  
A/Accession: T09059  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-1964 <ROW>  
A/Cross-references: UNIPROT:P31695; UNIPARC:UPI000016C7F1; EMBL:AF030001; NID:G2564945; F  
C/Genetics:  
A/Gene: notch4

A:Map position: 17  
 A:Introns: 22/1, 49/2, 148/1, 264/1, 305/1, 384/1, 436/1, 501/1, 539/1, 577/1, 618/1, 671/3, 1729/1, 1761/3  
 C:Superfamily: notch protein, ankyrin repeat homology; EGF homology  
 C:Keywords: receptor, signal transduction  
 F:514-545/Domain: EGF homology <EGF>

## Alignment Scores:

| Pred. No.:             | 2,22   | Length:       | 1964 |
|------------------------|--------|---------------|------|
| Score:                 | 101.00 | Matches:      | 64   |
| Percent Similarity:    | 28.5%  | Conservative: | 15   |
| Best Local Similarity: | 23.1%  | Mismatches:   | 96   |
| Query Match:           | 6.5%   | Indels:       | 102  |
| DB:                    | 2      | Gaps:         | 17   |

US-10-537-002-7 (1-786) x T09059 (1-1964)

```

QY 9 GACTGCTGTTCAGAGGCTTGCGGCTTCGTTCACTGATTGGGATTCGGGATCATTCG 68
DB 573 GTCybleuProglYpHeGLU-----GlyProHIScys 583
QY 69 TCCCACTTCAGTAGACCAAGTGAACCCAAAGACTTGTCAACAACACC-----CGT 119
DB 584 -----GluYsgluValaIaerGluCysleuSerAspProCysProValGly 598
QY 120 AACAGCTGTTTCAACTACAGAGGCTGTGGGCTTCGTGTCCGAGAGAGCTTGGCTT 179
DB 599 AlaSerCysleuAspIleuProglYalaIaerHeCysleuCyalaH----- 613
QY 180 CACCGAGTCCCGGGCTACTTAC-----CTGTGGGCTGCCAGCCAT 224
DB 614 -----ProglYpHeThrGlyIleuGluCysgluValaIaerProleuCythrProAsnMet 630
QY 225 GCTGACGAGAGTGCAGAGCCCTATATCTAGGCACTTCCTGGGTCATTTGGCTCT 284
DB 631 CysgluProglYgluIleuGluCysgluIleuGluIleuValaIaerProCys----- 646
QY 285 GGTATCATCTTTCGCTGAAATGATCCGACATTTGGACAGATGAGAGACTGCCAAGC 344
DB 647 -----LeuCythrProAspIleuSerProglY-----CysValPro 657
QY 345 CAACATGACATGACCTCCGGATCATGTTCTGAGTCTTGTGCAATTGC-- 401
DB 658 AlaIuAspAsnCythrProCysHISgluHIScysgluIleuSerleuCyValaIaer 677
QY 402 -----TGAGT-----GTCTGTCTT-----TGC 419
DB 678 GluIuYpHeThrGlyProglYcysgluIleuGluIleuGlyCysIleuSerThrProCys 697
QY 420 CAACATGCTGTGACTTAATCTTGATGTCACAGATCAATGATACACCGGATGGGTG 479
DB 698 AlaHISgluGly-----ThrCysHISProglYProserIleuYpHe 710
QY 480 GAT-----GCTGACAGCTGTTCAGACCGAGGTACACATTGG 515
DB 711 AsnCythrCysProAlaGlyIleuMetGlyleuThrCysSerGluIleuValaIaer 729
QY 516 TGGCGCTCTGTTCCGGGCTGGGCTGGAGCCCTACACATTAATGGGGGTGATAT 575
DB 730 CysHISser-----GlyProCysleuAsnIleuIleuSerCysSer 742
QY 576 GGCATGCGCTGCCGGG-----CTGGACACGAGAAACCAACTA 617
DB 743 Ile-----ArgProgluIleuYpHeSerCysThrCysleuProSerHISThrIleuArgHIScys 761
QY 618 CAAGCCGTTTCTTATCATGCTTCAGGCGCACAGTGT----- 653
DB 762 GluThrAlaValaIaerHIScysValaIaerAspCysleuAsnIleuGlyIleuThrCysValaIaer 781
QY 654 -----TGCCTACAGAGCTTGAGGCTTCAAGGCGACAGAC 686
DB 782 LysProglYpHeThrPhePheCysleuCyalaIleuGlypHeGluIleuHIS 798

```

## RESULT 21

S45306  
 notch 3 protein - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004  
 C:Accession: S45306  
 R:Laedel, M.; Dahlstrand, J.; Lendahl, U.

Mech. Dev. 46, 123-136, 1994

A:Title: The novel Notch homologue mouse Notch 3 lacks specific epidermal growth factor-1

A:Reference number: S45306; MUID:9500156; PMID:7918097

A:Accession: S45306

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-2318 &lt;LAR&gt;

A:Cross-references: UNIPROT:061982; UNIPARC:UPI000002930C; EMBL:X74760; NID:G483580; PID

C:Superfamily: notch protein, ankyrin repeat homology; EGF homology

F:163-195/Domain: EGF homology &lt;EGF1&gt;

F:474-505/Domain: EGF homology &lt;EGF2&gt;

F:854-885/Domain: EGF homology &lt;EGF2&gt;

F:1839-1871/Domain: ankyrin repeat homology &lt;AN1&gt;

F:1872-1904/Domain: ankyrin repeat homology &lt;AN1&gt;

F:1906-1938/Domain: ankyrin repeat homology &lt;AN3&gt;

F:1939-1971/Domain: ankyrin repeat homology &lt;AN4&gt;

F:1972-2004/Domain: ankyrin repeat homology &lt;AN5&gt;

## Alignment Scores:

| Pred. No.:             | 2,4    | Length:       | 2318 |
|------------------------|--------|---------------|------|
| Score:                 | 100.50 | Matches:      | 60   |
| Percent Similarity:    | 29.3%  | Conservative: | 22   |
| Best Local Similarity: | 21.4%  | Mismatches:   | 85   |
| Query Match:           | 6.4%   | Indels:       | 113  |
| DB:                    | 2      | Gaps:         | 17   |

US-10-537-002-7 (1-786) x S45306 (1-2318)

```

QY 3 GCGCGTGAAGCTGTGTCAGAGGCTTGCGGCTTCGTTCACTGATTGGGATTCGGGAT 62
DB 670 GluYleuSerCysValaIaerGly-----GluAsnIleuPhe 680
QY 63 CATTCCTGCCACCTGATGGA-----CCAGTGGAGACCCAGAGA 101
DB 681 HIScysleuCythrProglYserIleuProleuCybleuProAlaAsnIleuPro 699
QY 102 CTGTGACAAACACC-----CGTAAAGCTGTTTCAACTACAGAGGCTGTGGCGCTC 155
DB 700 CysAlaHISgluYpHeProCysSerHISgluValaIaerAspAlaIaerProglYpHeArgCys 719
QY 156 CTGTGTCCGAGAGAGCTGTGCTTACCGAGTCCGGGCTTACCTTCACTGCTGGGCT 215
DB 720 ValCys-----GluProglYpHeSer 726
QY 216 GCGAGCATGCTGACAGGAGCGGAGCCCTGATGATCTAGGACATTCGCT 266
DB 727 GlyProArgCysSerGluIleuAlaIaerProAspAlaCysgluSerGluProCysgluIleu 746
QY 267 -----GGGTCCCATTTGGCTCTGATTC-- 290
DB 747 GluYleuThrCysThrSerAspIleuIleuPheArgCysThrCysAlaIaerProglYpHeGlu 766
QY 291 ---CATTTTGCTGAAATGATCCG-----CATTTGACAGATGAGGACTC 335
DB 767 GluHISgluIleuValaIaerleuSerProCysThrProserleuCygluHISgluHIS 786
QY 336 TGGCAAGCCAAACAGACATGACTCCGGGATCATGTTCTGAGGCTTGTGCG 395
DB 787 CysgluSerAspProAspArg----- 793
QY 396 AATTCCTGAAGTGTCTGTTTCCCAACATGCTGTGACTTAATCTTGATGTCACAGAC 455
DB 794 -----LeuThrValaIaerCysSerCysPro 800
QY 456 TAACTGTACACCGGAGTGGGATGTCGACACTGTTACAGACAGCTA----- 506

```

```

Db 801 -----ProGlyTTPrgInGlyProArgCysGlnGlnAspValAspGlyCys 815
Qy 507 -----CACATTGGTGGGCTGTCTGTCGGGCTGGGCTGC 542
Db 816 AlaGlyAlaSerProCysGlyProHisGlyThrCysThrAsnLeuProGlyAsnPheArg 835
Qy 543 -----TGAGAGCCTACACATAATTGGGGGTGTGAT---GATTCGATGGCCCTG--- 587
Db 836 CysIleCysHisAlaArgGlyLysThrGlyProPheCysAspGlnAspIleAspAspCysAsp 855
Qy 588 CCGGGGAGCCT-----GGACACAGAGAAACCACTACAAAGCCGTTTCTTATCA 635
Db 856 ProAsnProCysLeuHisGlyGlySerCysGlnAspGlyValGlySerPheSerCysSer 875
Qy 636 TGCCTC-----AGGCCACAGTGT-----TGCCTTACAGCC 665
Db 876 CysLeuAspArgLysPheHisGlyProArgCysValAspAspValAspGlyCysLeuSerSer 895

RESULT 22
S45306
notch 3 protein - mouse
C/Species: Mus musculus (house mouse)
C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #ext_change 09-Jul-2004
C/Accession: S45306
R/Lardelli, M.; Dahlstrand, J.; Lendahl, U.
Mech. Dev. 46, 123-136, 1994
A/Title: The novel Notch homologue mouse Notch 3 lacks specific epidermal growth factor-
A/Reference number: S45306; MUID:95001556; PMID:7918097
A/Accession: S45306
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-2318 <LNR>
A/Cross-references: UNIPROT:061982; UNIPARC:UPI00002930C; EMBL:X74760; NID:9483580; PTD
C/Suprafamily: notch protein; ankyrin repeat homology; EGF homology
F/163-195/Domain: EGF homology <EGF1>
F/474-505/Domain: EGF homology <EGF>
F/854-885/Domain: EGF homology <EGF2>
F/1839-1871/Domain: ankyrin repeat homology <AN1>
F/1872-1904/Domain: ankyrin repeat homology <AN2>
F/1906-1938/Domain: ankyrin repeat homology <AN3>
F/1939-1971/Domain: ankyrin repeat homology <AN4>
F/1972-2004/Domain: ankyrin repeat homology <AN5>

Alignment Scores:
Pred. No.: 2.4 Length: 2318
Score: 100.50 Matches: 64
Percent Similarity: 28.4% Conservative: 19
Best Local Similarity: 21.9% Mismatches: 64
Query Match: 6.4% Indels: 145
DB: 2 Gaps: 17

US-10-537-002-7 (1-786) x S45306 (1-2318)
Qy 702 GTTGAACCCAAAGCCAGTCTGCTGCTGTAAGCTCCAGGCTTGTAGCAACATGTGGCC 643
Db 1167 LeuGlyProSerLeuAspSerGlyValGlnCys----- 1177
Qy 642 TAGGCGATGATGAAGAACGGCTTT-----GTAGTTGGTTTCTTGTGTCGACAGCC 592
Db 1178 -----LeuHisAsnGlyThrCysValAspLeuValGlyLysPheArgCysAsnGly 1194
Qy 591 CCGGAGCGGATGACATCATCACACCCCAATTAGTGTAGGCTTCACGACCAAGCC 532
Db 1195 ProProGlyLysThrGlyLeuHis-----CysGlnAlaAspIleAsn----- 1208
Qy 531 CACGAACAGAGCGGACCAATGTACCTGTGTGAACAGTGTGACCATCCACCAT 472
Db 1209 -----GlnCysArgProGly-----AlaCysHis 1216
Qy 471 GCCGATGTACATGTAGCTGTGTGACATCCAGAAATTAGTCAACGATGTGGCAAC 412
Db 1217 AlaAla----- 1218

```

```

Qy 411 AGACATCCAGCAATTTGCAAAAGACCTGAGACATGAACATGATCCGGAGGTGAC--- 355
Db 1219 ---HisThrArgAspCysLeuGln-----AspProGlyLysHisPhe 1231
Qy 354 -----TGTCAATTTGGCTTTTGGCAGAGTCTTCATGC----- 323
Db 1233 ArgCysValCysHisProGlyPhe---ThrGlyProArgCysGlnIleAlaLeuSerPro 1251
Qy 322 -----TGCACATCCGATGTCATTCAGGGCAAAAGATGATACAGAGAGCC 277
Db 1251 YeGlySerGlnProCysGln----- 1257
Qy 276 AATGGACACCCAGAGAGATGCTTACATCATCAGGGCTC----- 239
Db 1258 -----HisGlyLysGlnCysArgHisSerLeuGlyArgGlyLysLeuThrPheThrC 1276
Qy 238 -----GCATGCTCTCAGCATG----- 221
Db 1276 YeHisCysValProProPheThrGlyLeuArgCysGlnValAlaArgSerCysArgG 1296
Qy 220 ---CTGGCAGCCCGCAGCAGGTTAAGTAAAGTCCGCGTGAAGCCAGACTCTCTG 163
Db 1296 LysLeuGlnCysProValGlyLysProCys-----GlnGlnThrAlaArgG 1311
Qy 162 GACACAGAGGCGGACAGCCCTGGTAAAGTGAACAGCTGTACGGGT-----TGT- 110
Db 1311 LysProArgCysAlaCysProProGly-----LeuSerGlyProSerCysA 1326
Qy 109 -----TGTACAGTCTTGGTGTCTTCACATGTCATGCA 76
Db 1326 rgValSerArgAlaSerProSerArgLysAlaThrAsnAlaSerCysAlaSerAlaProCysL 1346
Qy 75 GGTGGACGATGATGATGCGCCGATCCATCCATCAGTGA----- 38
Db 1346 euHisGlyGlySerCysLeuProValGlnSerValProPheArgCysValAspAlaP 1366
Qy 37 -----CCAGGACCCCAAGCCCT 20
Db 1366 roGlyTTPrgLysGlyProArgCysGlnThrPro 1376

RESULT 23
PC2061
genome polypeptide N2 - hepatitis C virus
N/Contains: envelope protein B1; nonstructural protein B2/NS1
C/Species: hepatitis C virus
C/Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #ext_change 09-Jul-2004
C/Accession: PC2061
R/Li, J.S.; Vitvitski, L.; Tong, S.P.; Trepo, C.
Biochem. Biophys. Res. Commun. 199, 1474-1481, 1994
A/Title: Identification of the third major genotype of hepatitis C virus in France.
A/Reference number: PC2060; MUID:94197744; PMID:8147893
A/Accession: PC2061
A/Molecule type: mRNA
A/Residues: 1-411 <LNR>
A/Cross-references: UNIPROT:081813; UNIPARC:UPI00000900B; GB:U12355; NID:9410169; PTD:
C/Suprafamily: hepatitis C virus genome polypeptide; glycoprotein; nonstructural protein;
C/Keywords: ATP; capsid protein; envelope protein; glycoprotein; nonstructural protein;
F/192-383/Product: envelope protein B1 #status predicted <SPE>
F/384-411/Product: nonstructural protein B2/NS1 #status predicted <NP>
F/196,209,234,305,325/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 3.04 Length: 411
Score: 100.00 Matches: 57
Percent Similarity: 36.8% Conservative: 34
Best Local Similarity: 23.1% Mismatches: 102
Query Match: 6.4% Indels: 55
DB: 2 Gaps: 11

US-10-537-002-7 (1-786) x PC2061 (1-411)
Qy 26 TGGGTTCCGTGTTTCATGATTTGGATTCGGGATTCATTCGACCTGACATGAGCC 85

```

```

Db      93  TPrLaGIyTrPLeu-----LeuSerProArgLySer-----ArgProSerTrpGly 108
QY      86  AGTGAAGACCCCAAGACTTGTAACAACACCCGGTAACACTGTTTCACTACCAAGGGGC 145
Db      109  ProAnaMPProArg-----ArgArgSerArgAna 118
QY      146  TGTGGCGCTCGTGTTCAGAGAG-AGCTGTGGCTTCACAGAGTGGCGGGCTACTTACAC 204
Db      119  LeuGIyLySValIIlaSPThrLeuThrCySgLyPheAlaSPLeuMetGlyTrIlePro 138
QY      205  CTGCTGGGGCTGCCA-----GCCATGCTGCAGGCACTGGAGCCCTG 246
Db      139  LeuValGIyAlaProValGIyLyAlaAlaArgAlaLeuAlaHISgLyValArgAlaLeu 158
QY      247  ATGATC---GTAGGATCGTCCGGGTGGCATT---GGCTCTGTGTATCCATCTTTGGCC 300
Db      159  GluAerGIyIIlaSPnPheAlaThrGIyAlaLeuProGIyCySerPheSerIlePheLeu 178
QY      301  CTG-----AAATGCATCCGATTTGGCAGACATGAGAGACTTGCACCAAGCCAAACATG 351
Db      179  LeuAlaLeuPheSerCySLeu-----IleHISProAlaAlaSerIleuGIuTrp 194
QY      332  ACACTGACCTCCGGATCATATGTTCAATGTTCTTGGCAATTGCTGAGAGTCT 411
Db      195  ArgAnaThrSerGIy---LeuTyxIleLeuThrAsnAspCySerAsnSerSerIleVal 213
QY      412  GTGTTGGCAACATGCTGGTGACTTAATTCTC-----TGATGTCCACAGCT 456
Db      214  TyrGIuAlaAspAspValIIleuHISThrProGIyCySValProCySValGIaSPgLy 233
QY      457  AACATGTACACCGGATGGGTGGATGTGCAGACTGTTGCAGACCAAGTAC----- 507
Db      224  AsnThrSerThrCySgTrpThrProValThrProThrValAlaValArgTrpValGIyAla 253
QY      508  -----ACATTGGTGGCG 519
Db      254  ThrThrAlaSerIleArgSerHISValaSPLeuLeuValGIyAlaAlaIleThrMetCySer 273
QY      520  GCTGTGTTGCGGGCTGGGTGGTGGAGGCTCACACTAATTGGGGGTGTATGATGTGC 579
Db      274  AlaLeuTyxValGIyAlaSPMetCySgLyAlaValPheLeuValGIyGIaIlaPheThrPhe 293
QY      580  ATGCGCTGCGGGGCTGGCAGACCAAGAAACCACTACAAAGCCGTTCTTATCATGGCC 639
Db      294  ArgProArgArgHISgIlnThrValGIlnThrCySAsnCySerLeuTyxProGIyHISLeu 313
QY      640  TCAGGCCACAGTGTGCCTTAC 660
Db      314  SerGIyHISArgMetAlaTrp 320

RESULT 24
S41288
Genome polypeptide - hepatitis C virus (fragment)
NContains: core protein; envelope protein; NS1 protein
CSpecies: hepatitis C virus
CDate: 06-Jan-1995 #sequence_revision 26-Jul-1996 #text_change 31-Dec-2004
CAccession: S41288
RSeelig, R.
submitted to the EMBL Data Library, December 1993
AReference number: S41288
AAccession: S41288
AMolecule type: genomic RNA
ARebids: 1-492 <SRES>
ACross-references: UNIPROT:068870; UNIPARC:UPI0000178522; EMBL:X76918
CKeywords: capsid protein; core protein; envelope protein; nonstructural protein; polyP
F1-191/RProduct: core protein #status predicted <CON>
F1192-372/RProduct: envelope protein #status predicted <ENV>
F1373-492/RProduct: NS1 protein (fragment) #status predicted <NS1>

Alignment Scores:
Pred. No.: 3 Length: 492
Score: 100.00 Matches: 56
Percent Similarity: 36.8% Conservative: 35

```

```

Best Local Similarity: 22.7% Mismatches: 102
Query Match: 6.4% Indels: 55
Db: 2 Gaps: 11

US-10-537-002-7 (1-786) x S41288 (1-492)

QY      26  TGGGGTGGTGGTTTCACTGATTTGGAGATGAGGATGCTGTCGACCTGATGAGCC 85
Db      93  TPrLaGIyTrPLeu-----LeuSerProArgLySer-----ArgProSerTrpGly 108
QY      86  AGTGAAGACCCCAAGACTTGTAACAACACCCGGTAACACTGTTTCACTACCAAGGGGC 145
Db      109  ProAnaMPProArg-----ArgArgSerArgAna 118
QY      146  TGTGGCGCTCGTGTTCAGAGAG-AGCTGTGGCTTCACAGAGTGGCGGGCTACTTACAC 204
Db      119  LeuGIyLySValIIlaSPThrLeuThrCySgLyPheAlaSPLeuMetGlyTrIlePro 138
QY      205  CTGCTGGGGCTGCCA-----GCCATGCTGCAGGCACTGGAGCCCTG 246
Db      139  LeuValGIyAlaProValGIyLyAlaAlaArgAlaLeuAlaHISgLyValArgAlaLeu 158
QY      247  ATGATC---GTAGGATCGTCCGGGTGGCATT---GGCTCTGTGTATCCATCTTTGGCC 300
Db      159  GluAerGIyIIlaSPnPheAlaThrGIyAlaLeuProGIyCySerPheSerIlePheLeu 178
QY      301  CTG-----AAATGCATCCGATTTGGCAGACATGAGAGACTTGCACCAAGCCAAACATG 351
Db      179  LeuAlaLeuPheSerCySLeu-----IleHISProAlaAlaSerIleuGIuTrp 194
QY      332  ACACTGACCTCCGGATCATATGTTCAATGTTCTTGGCAATTGCTGAGAGTCT 411
Db      195  ArgAnaThrSerGIy---LeuTyxIleLeuThrAsnAspCySerAsnSerSerIleVal 213
QY      412  GTGTTGGCAACATGCTGGTGACTTAATTCTC-----TGATGTCCACAGCT 456
Db      214  TyrGIuAlaAspAspValIIleuHISThrProGIyCySValProCySValGIaSPgLy 233
QY      457  AACATGTACACCGGATGGGTGGATGTGCAGACTGTTGCAGACCAAGTAC----- 507
Db      224  AsnThrSerThrCySgTrpThrProValThrProThrValAlaValArgTrpValGIyAla 253
QY      508  -----ACATTGGTGGCG 519
Db      254  ThrThrAlaSerIleArgSerHISValaSPLeuLeuValGIyAlaAlaIleThrMetCySer 273
QY      520  GCTGTGTTGCGGGCTGGGTGGTGGAGGCTCACACTAATTGGGGGTGTATGATGTGC 579
Db      274  AlaLeuTyxValGIyAlaSPMetCySgLyAlaValPheLeuValGIyGIaIlaPheThrPhe 293
QY      580  ATGCGCTGCGGGGCTGGCAGACCAAGAAACCACTACAAAGCCGTTCTTATCATGGCC 639
Db      294  ArgProArgArgHISgIlnThrValGIlnThrCySAsnCySerLeuTyxProGIyHISLeu 313
QY      640  TCAGGCCACAGTGTGCCTTAC 660
Db      314  SerGIyHISArgMetAlaTrp 320

RESULT 25
T01437
Hypothetical protein R34001.1 - human
CSpecies: Homo sapiens (man)
CDate: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
CAccession: T01437
RFlamerdin, J.E.; McCreedy, P.M.; Skowronski, E.; Adamson, A.W.; Burkhat-Schultz, K.; Ge
P.; Christensen, M.; Georgescu, A.; Avila, J.; Liu, S.; Bruce, R.; Quan, G.; Krommiller,
M.; Coefield, J.; Amico-Keller, G.; Lucac, S.; Duarte, S.; Olsen, A.O.; Carrano, A.V.
submitted to the EMBL Data Library, February 1998
ADescription: Sequence analysis of a 1 Mb region in 19q13.1.
AReference number: Z14330
AAccession: T01437
AStatus: preliminary; translated from GB/EMBL/DBJ
AMolecule type: DNA

```



A/Residues: 1-1007 <LIM>  
 A/Cross-references: UNIPROT:P043379; UNIPARC:UPI000006CC3; EMBL:AC004444; NID:G2887496;  
 C/Genetics:  
 A/Introns: 82/2; 113/1; 155/1; 177/3; 218/2; 243/3; 332/2; 398/2; 416/1; 487/3; 513/2; 5  
 Alignment Scores:  
 Pred. No.: 2.82 Length: 1007  
 Score: 100.00 Matches: 45  
 Percent Similarity: 31.3% Conservative: 12  
 Best Local Similarity: 24.7% Mismatches: 45  
 Query Match: 6.4% Indels: 81  
 DB: 2 Gaps: 11  
 US-10-537-002-7 (1-786) x T01437 (1-1007)

```

OY 75 CTGATGACGACAGT-----GAGCACCAGACTGTGACA 110
DB 842 LeuHISGLYSerAlaPheArgProSerLeuProAlaProGluSerProGlyLeuProAla 861
OY 111 CAACCCCGTAACAGCTGTTTCACTACACGAGGGCTGT-----GGCCTCTGTGTCCGAG 166
DB 862 HIsProSerAsnPro-----GlnLeuProGluAlaArgProGlyLeuProGlyLeuThr 879
OY 167 AGAGCTGTGCTTCAACGAGCGGGGCTACTTCACTGCGGGCTGCGACCATGC 226
DB 880 AlaSerLeuLeuGluProThrSerGlyTyrGlyThrSerCysThrGlyCysArgProPro 899
OY 227 TGCAGCAG-----TGCAGCCCTGATGATCGTAGCATGCTCTG 268
DB 900 SerLysLysProSerThrPheThrValCys-----Trp 910
OY 269 GTGGCATGCGCTCTGCTGATTCATCT-----TTGCCCTGA 304
DB 911 SerProValAlaArgTyrThrProGlySerSerArgHISGlyLeuSerTrpSerProPro 930
OY 305 AATGATCCGATGCGCATGAGAG-----ACTGCGCAAG 343
DB 931 SerCysGlySerThrAlaSerTrpArgLeuAlaIleTrpGlyLeuValTrpGln 950
OY 344 CCACATGACACTGACCTCCGGATCAGTTCATGTCAGGCTTGTGCAATTGCTG 403
DB 951 ProArg----- 952
OY 404 GAGTGTCTGTGTGCA-----ACATGCTGTGACTA 436
DB 953 -----LeuCysArgAlaGlnAspProArgProHISArgArgCysThrProTrp----- 968
OY 437 ACTTGTGATGTCCACAGCTAACATGTACACGGCATG-----GTGGGATGG 484
DB 969 -----ProAlaGlnThrCysArgProCysTrpAsnThrThrArgSerCysTrp 984
OY 485 TGCAGA 490
DB 985 CysArg 986
RESULT 26
S74569
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 2 - Synecchocystis sp. (strain PCC 68
N/Alternate names: hypothetical protein ndhB
C/Species: Synecchocystis sp.
A/Variety: PCC 6803
C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #ext_change 09-Jul-2004
C/Accession: S74569; A39366
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asami, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis
S.
A/Reference number: S74322; MUID:97061201; PMID:8905231
A/Accession: S74569
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-521 <KAN>

```

A/Cross-references: UNIPROT:P2714; UNIPARC:UPI000013065D; EMBL:D90900; GB:AB001339; NID:  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 R/Ogawa, T.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 4275-4279, 1991  
 A/Title: A gene homologous to the subunit-2 gene of NADH dehydrogenase is essential to tr  
 A/Reference number: A39366; MUID:91239550; PMID:1903537  
 A/Accession: A39366  
 A/Molecule type: DNA  
 A/Residues: 1-411, 'R', 413-521 <OG>  
 A/Cross-references: UNIPARC:UPI00000BAED0; GB:D90288; NID:G217096; PDB:BAI4330.1; PID:6  
 C/Genetics:  
 A/Gene: ndhB  
 C/Superfamily: NADH dehydrogenase (ubiquinone) chain 2  
 C/Keywords: membrane-associated complex; NAD; oxidoreductase  
 Alignment Scores:  
 Pred. No.: 3.6 Length: 521  
 Score: 99.00 Matches: 69  
 Percent Similarity: 34.6% Conservative: 33  
 Best Local Similarity: 23.4% Mismatches: 75  
 Query Match: 6.3% Indels: 118  
 DB: 2 Gaps: 17  
 US-10-537-002-7 (1-786) x S74569 (1-521)

```

OY 4 GCCGTGACTGCTGTCCAGGGCTTGCTGTTGACTGAT-----GGG 51
DB 205 AlaAspThrValGlyGlnSerLeuValAlaIleValAlaValAlaValAlaValAlaVal 224
OY 52 ATGGGCGCATCATGCTGCCACCTGATGACGACGAGGACCCAGACTGTGACAC 111
DB 225 IleAlaPheLysIleSerAlaValProPheHISGlnTrp-----ThrProAspValArgIle 243
OY 112 AACCCGTAACGCTGTTTCACTACACAGGGGCTGTG----- 150
DB 244 GlySerProThrProValAlaPheLeuSerValGlySerLysAlaAlaGlyPheAla 263
OY 151 -----CGTCTCTGTGTCGAGAGAGCTTGCTTCCAC-----GAGTCCGGGGCTAC 198
DB 264 ValAlaIleArgLeuValThrAlaPheGlyIleIleHisPheLysIleValIle 283
OY 199 TTCACCTGTGCGGCTGCCAGCCATG-----CTGCAGGAGTGCAGCCCTG----- 246
DB 284 PheThrAlaLeuAlaValLeuSerMetValLeuGlyAsnValValAlaLeuAlaGlnThr 303
OY 247 -----ATGATCGTAGGC 258
DB 304 SerMetLysArgMetLeuAlaTyrSerSerIleGlyAlaGlyPheValMetIleGly 323
OY 259 ATCGTCCGTGGGCTG-----ATGGCCCTC 282
DB 324 LeuValAlaGlySerGlnAspGlyTyrAlaSerMetValPheTyrMetLeuIleTyrLeu 343
OY 283 CTGTATCCATCTTTCCTGGAATGATC-----CGCATGCGACATG 327
DB 344 PheMetLeuLeuGlyAlaPheSerCysIleIleLeuPheThrLeuArgThrGlySerAsp 363
OY 328 GAGACTGTGCCAAGCCACATG-----ACATGACTCCGGATC----- 369
DB 364 GlnIleSerAspLysAlaGlyLeuTyrHisLysAspProLeuLeuThrLeuIleLeuSer 383
OY 370 -----ATGTTATGTCGACGGTCTTGTGTAATGCTGAGTGTGTGTTGCCAAC 423
DB 384 IleCysLeuLeuSerLeuGlyGlyIleProProLeuAlaGly-----PheHISGlyLys 401
OY 424 ATGTGTGTGACTTACTTGTGATGTCCACGCTAACATGTACACCGGATGGGTGGATG 483
DB 402 IleTyrIle-----PheTrp----- 406
OY 484 GTGCAGACTGTTCAGACCAAGTACATTTGTGCGGCTGTGTTGGGTGGTGGCT 543
DB 407 -----AlaGlyTyrPheIleSer 411

```

QY 544 GGA-----GGCTCACTAATGCGGCTG----- 570  
 DB 412 GlyLeuTYRGLyLeuValLeuLeuGLyLeuValThrSerValValSerIleTYRtyr 431  
 QY 571 -----ATGATGTGCATCGCTGCCGGGCGCTGCACACAGAGAA----- 609  
 DB 432 IleArgValValIleuSerMetValVallys-----GluProGlnGluMetSer 447  
 QY 610 -----ACCACTACAAAGCCGTTTTCATCATGCTCAGGC 645  
 DB 448 GluValIleLysAsnTYRProAlaIleLysTRPAsnLeuProGly 462

## RESULT 27

A47282  
 Calcium-binding protein calphotin - fruit fly (Drosophila melanogaster).

C.Species: Drosophila melanogaster  
 C.Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C.Accession: A47282  
 R.Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 1531-1535, 1993  
 A.Title: Calphotin: a Drosophila photoreceptor cell calcium-binding protein.  
 A.Reference number: A47282; MUID:93165729; PMID:8094559  
 A.Accession: A47282  
 A.Status: preliminary  
 A.Molecule type: nucleic acid  
 A.Residues: 1-865 <MAR>  
 A.Cross-references: UNIPROT:P02910; UNIPARC:UPI00001282D3; GB:L02111; NID:g157031; PIDN:  
 A.Experimental source: photoreceptor cells  
 A.Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBI:P124956)  
 C.Genetics:  
 A.Gene: FlyBase:Cpn  
 A.Cross-references: FlyBase:FBgn010218  
 C.Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C.Keywords: calcium binding

## Alignment Scores:

Pred. No.: 3,45 Length: 865  
 Score: 99.00 Matches: 60  
 Percent Similarity: 35.8% Conservative: 33  
 Best Local Similarity: 23.1% Mismatches: 101  
 Query Match: 6.3% Indels: 66  
 DB: 2 Gaps: 12

US-10-537-002-7 (1-786) x A47282 (1-865)

QY 695 CGAAGCCAGTCTGCGCTTGAAGCT---CGAGCTTGAGCAACACTGCGCTGAG 639  
 DB 362 ProAlaProValAlaThrThrProValProAlaThrLeuAlaValThrAspProAsp 381  
 QY 638 GCATGATAAGAAACGCTTGTAGTTCTTCTGTGTCAGCGCCCGCGAGCGGATG 579  
 DB 382 -----ValThrAlaSerAlaValPro---GluLeuPro 391  
 QY 578 CACATCATCACCC-----CCAAATTAGTGTAGG 549  
 DB 392 ProValIleAlaProSerProValProSerAlaValAlaGluThrProValAspLeuAla 411  
 QY 548 CCTCAGACGACCGAGCC---ACGAACAGACCGCAGCAAGTGTACTGCTGTAAGA 492  
 DB 412 ProProValLeuProProValAlaAlaGluProValProAlaValAlaGluGluThr 431  
 QY 491 GTCTGCACATCCACCC-----ATCCGGTGTACTGTTGATGCTGTGACATCCAGAG 438  
 DB 432 ProGluThrProAlaProAlaSerAlaProValThrIleAlaAlaLeuAspIleProGlu 451  
 QY 437 TTAGTCACGACGATGTTGGCAACACAGCACTCCAGCA----- 399  
 DB 452 ValAlaProValIleAlaAlaProSerAspAlaProAlaGluAlaProSerAlaAlaAla 471  
 QY 398 ---ATTGCAACAAAGCTTGACAAATGAACATGATCCGAGAGTCAGTGTGCT 342  
 DB 472 ProIleValSerThrProThrThrAlaSerValProGluThrThrAlaProProAla 491

QY 341 TTGGCAGAGTCCTCCATGCTGCGCAATGCGGATTCACAGGCAAGATGATACAGG 282  
 DB 432 AlaValProThrGluProIleAspValSerValLeuSerGluAlaAlaIleGluThr--- 510  
 QY 281 AGGCCAATGGACCCGACGACGATGCTGCATCATCAGAGGCTGCGACTGCTGCATG 222  
 DB 511 ---ProValAlaPro-----ProValGluValThrThrGluValAlaValAlaAsp 526  
 QY 221 GCTGCACACCCCGACGAGGTTGAG-----TACGCCCGGACATCG 183  
 DB 527 ValAlaProProGluAlaAlaAlaAspLeuIleIleGluProValGluProProAlaPro 546  
 QY 182 GTGAAGCCAGAGCTCTCTCGACACGAGCGCCACAGCCCTGTGATGTTGAAAACAGCT 123  
 DB 547 IleProAspLeuLeuGlnGlnThr-----ThrSer 556  
 QY 122 GTTACGGGGTGTGTTACAAAGTCTTGCTGCTCCACTGCTCATGCTGACAGTGCACATG 63  
 DB 557 ValProAlaVal-----GluAlaAlaGlu 564  
 QY 62 ATGCCCGAATCCCAATGATGTAACCAAGAACCCCAAGCCCTGACAGGCAATCAGGC 3  
 DB 565 SerThrSerProIleProGluThrSerLeuProProAlaGlnGluAlaAlaSer 584

## RESULT 28

S57246  
 ventral nervous system defective protein - fruit fly (Drosophila melanogaster)

N.Alternate names: NK-2 homeotic protein

C.Species: Drosophila melanogaster

C.Date: 28-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 05-Oct-2004

C.Accession: S57246; B33976

R.Jimenez, F.; Martin-Morris, L.E.; Velasco, L.; Chu, H.; Sierra, J.; Rosen, D.R.; White,

EMBO J. 14, 3487-3495, 1995

A.Title: vnd, a gene required for early neurogenesis of Drosophila, encodes a homeodoma

A.Reference number: S57246; MUID:95354667; PMID:7628450

A.Accession: S57246

A.Molecule type: mRNA

A.Residues: 1-722 <JIM>

A.Cross-references: UNIPROT:P22808; UNIPARC:UPI000012CA4F; EMBL:X87141; NID:g1045047; PII

Proc. Natl. Acad. Sci. U.S.A. 86, 7716-7720, 1989

A.Title: Drosophila NK-homeobox genes.

A.Reference number: A33976; MUID:90046666; PMID:2573058

A.Accession: B33976

A.Status: preliminary

A.Molecule type: DNA

A.Residues: 476-631, 'VG' <JIM>

A.Cross-references: UNIPARC:UPI000016BC38; GB:M27290; NID:g157635; PIDN:AAA28617.1; PID:5

C.Keywords: DNA binding; homeobox; nucleus; transcription regulation

F/545-601/Domain: homeobox homology <HOK>

## Alignment Scores:

Pred. No.: 3,85 Length: 722  
 Score: 98.50 Matches: 56  
 Percent Similarity: 32.5% Conservative: 30  
 Best Local Similarity: 21.1% Mismatches: 78  
 Query Match: 6.3% Indels: 101  
 DB: 2 Gaps: 15

US-10-537-002-7 (1-786) x S57246 (1-722)

QY 588 GAGGCGATGCAATCATCACACCCCAATTAGTGTAGGCTCCAGCGACCCAGCCAG 529  
 DB 275 AlaGlyGluAlaAspHisHisSerThrThr-----GluHisHisAlaProProSerHis 292  
 QY 528 GAACAGCGCGCACCAATGTGTACTGATCTGAACACTGTGCACCAATCCCAAGCCAGCC 469  
 DB 293 ProGlnGlnGln-----HisProHisHisGln 301  
 QY 468 GGTGTACATGTTAGCTGTGACATTCAGAA----- 439





1668 IlegluthrSerThrAlaProSerLeuGluGlyAlaProGlyGluThrSerGluThrSer 1707  
QY 203 -----GTGAAGTACCCCGGCACTGGTGAAGCCAGAGCTCTTGGACA 159  
DB 1708 ValSerIyValIleMet-SerSerProGlySerAlaSerSerSerIyS- 1724  
QY 158 CAGAGCCCGCAGCCCGCTGGT-----AGTTGAAGAAG-----CTGTTACG 117  
DB 1725 ArgAlaSerThrLeuProAlaThrThrLeuProSerLeuGlyAlaSerValLeuSe 1744  
QY 116 GGGTTGTGTACAACTCTTGGG-----TGCTC 90  
DB 1744 rProThrAlaThrSerSerGlyIyAspSerHisIleSerProValSerAlaAlaCysSe 1764  
QY 89 CACTGGTCCATGACAGTGGCAGCAATGATGCCCAATCCCAATGATGAACCAAC 30  
DB 1764 ThrGlyThrThr-----ThrProGlnAlaSerGluIyLeuProSerIy 1779  
QY 29 CCGAAGCCCT 20  
DB 1779 GlyGlyPro 1782

RESULT 32  
T09059  
notch4 - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C/Accession: T09059  
R/Rowen, L.; Mbehatas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Laaky, S.; Loretz, C.; Sc  
submitted to the EMBL Data Library, October 1997  
A/Description: Sequence of the mouse major histocompatibility locus class III region.  
A/Reference number: Z16543  
A/Accession: T09059  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-1964 <ROM>  
A/Cross-references: UNIPROT:P31695; UNIPARC:UPI000016C7F1; EMBL:AF030001; NID:92564945;  
C/Genetics:  
A/Gene: notch4  
A/Map position: 17  
A/Intons: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 67  
1679/3; 1729/1; 1761/3  
C/Superfamily: notch protein, ankyrin repeat homology; EGF homology  
C/Keywords: receptor, signal transduction  
P/514-545/Domain: EGF homology <EGF>

Alignment Scores:  
Pred. No.: 4.27 Length: 1964  
Score: 97.50 Matches: 60  
Percent Similarity: 31.84 Conservative: 17  
Best Local Similarity: 24.84 Mismatches: 75  
Query Match: 6.24 Indels: 91  
DB: 2 Gaps: 16

US-10-537-002-7 (1-786) x T09059 (1-1964)

QY 763 GATAGATTTGATCTGCTCTCTG-----TGCGGCACTTCATGATATCT 716  
DB 1004 AspGluCysLeuAspArgProCysHisProSerGlyThrAlaAlaCysHisSerLeuAla 1023  
QY 715 TCTTGTCTTTGGTGTGGAACCAAGCCAGTCTGGCCCTTGAAGCCTCCAGGCTTGAGG 656  
DB 1024 AsnAlaPheIyCys-----GlnCysLeuPro-----GlyHisThrGly 1036  
QY 655 CAACACTGGCGCTTGAAGCATGAT-----AAGAAACGCTTTGATG 614  
DB 1037 GlnAlaGCGGluValAlaGluMetAspLeuCysGlnSerGlnProCysSerAlaGlyIySer 1056  
QY 613 TGGTTCTCTTCTGATGCGCAGCCCGGCGCAGCCAGTGC----- 578  
DB 1057 CysGluIleThrThrGlyProProProGlyPheThrCysHisCysValProGlySerGlu 1076  
QY 577 AATATATGACACACCCCAATTGATGAGGCTTCAGCA----- 539

1077 GlyProThrCysSerHisIyAlaLeuSerCysGlyIleHisIleCysHisAlaSerGly 1096  
QY 538 -----CCAGCCCAACAAACAGAGCCGCAACAAATGTACTGCTGCAAGT-- 490  
DB 1097 LeuCysLeuProSerProGlySerProProLeuCysValaCys-LeuSerGlyPhe 1116  
QY 489 -----CTGACACATCCCAACCCGCGGTGTACATGTTAGCTGGACA 446  
DB 1116 eGlyGlyProAspCysLeuThrProProAlaProProGly-----CysGlyPhe 1132  
QY 445 TCCAGAAATTATGTCACACAGATGTTGGAAACACAGACATCCAGCAATTCAGCAAGAC 386  
DB 1132 oProSerProCys-----LeuHisAlaGlyThrCysThrGluThr 1145  
QY 385 CTGAGACATGAACATGATCCCGAGAGTCACTGATGTCATGTTGGCTTTGGCAGAGTCTCC 327  
DB 1145 r---ProGlyLeuGlnAsnProGlyPheGlnCysThrCys-----ProPhe 1159  
QY 326 A----- 326  
DB 1159 oAspSerProGlyProArgCysGlnArgProGlyAlaSerGlyCysGluIyArgGlyGly 1179  
QY 325 -----TGCTGCCAATGCGATGATTTTCAAGGCGCAAGATGATACAGAGGCGCAATG 273  
DB 1179 yAspGlyThrCysAspAlaGlyCys--SerIy-----ProGlyIyAspTr 1194  
QY 272 GCACCCAGACAGATGCTTACATGATATGACAGGCTGCACTGCTGAGAGATGGCTGGCAGC 213  
DB 1194 PAspGlyIyAspCys-----SerLeuGlyAlaProAspProTrIyGlyCys 1210  
QY 212 CCGA 209  
DB 1210 sPro 1211

RESULT 33  
T14271  
Doc4 protein, stress-induced - mouse  
N/Alternate names: odc protein homolog  
C/Species: Mus musculus (house mouse)  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: T14271  
R/Wang, X.Z.; Kuroda, M.; Sok, J.; Batchvarova, N.; Kimmel, R.; Chung, P.; Zinzner, H.;  
EMBO J. 17, 3619-3630, 1998  
A/Title: Identification of novel stress-induced genes downstream of chop.  
A/Reference number: Z17951; MUID:98315054; PMID:9649432  
A/Accession: T14271  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-2825 <MAN>  
A/Cross-references: UNIPROT:O70465; UNIPARC:UPI00000294B0; EMBL:AF059485; NID:93170614;  
C/Genetics:  
A/Gene: Doc4

Alignment Scores:  
Pred. No.: 4.15 Length: 2825  
Score: 97.50 Matches: 57  
Percent Similarity: 29.78 Conservative: 9  
Best Local Similarity: 25.74 Mismatches: 69  
Query Match: 6.24 Indels: 87  
DB: 2 Gaps: 16

US-10-537-002-7 (1-786) x T14271 (1-2825)

QY 70 GCAACTGATGAGCACTGAGACCAAGCTTGACAAACACCCGCTACAGCTGT 129  
DB 766 AlathrCysLeuAspGlnCysSerGlyHisGlyThrPhe-----LeuProAspThrGlyLeu 784  
QY 130 TTCAACTACACAGGAGCTGTGCGCTCC-----TGTCCTCGAAG 168  
DB 785 CysAsnCysAspProSerTrpThrGlyHisAspCysSerIleGluIleCysAlaIleAsp 804  
QY 169 AGCTCTGCTTACCGAGTGC-----CGGGGCTAATTACCCCTGC-----TGGGGCTG 216

```
Db      805 CysGlyLysIleGlyValCysValGlyLeu-ThrCysArgCysGlyAspGlyTyrMetC1 824
Qy      217 CCAAGCATGC---TGCAGGCAAGTGCAGAGCCCTGATGATGTAGGCAT-----CGTC 264
Db      824 ValAlaIcysAspArgIleAlaIcysIleHis-ProArgCysAlaGlyIleGlyThrCysArgG 844
Qy      265 CTGGG---TGCATTTGGGCTCCGGT-----ATGCATCTTTGGC 300
Db      844 spGlyLysCysGlyCysThrProGlyTyrAsnGlyLysIleCysThrIleGlyGlyCysP 864
Qy      301 CTGAATATGCATCCGATTTGGAGATGAGAGACTGTGCCAAGCCAAACATGACATGAC 360
Db      864 ro-----GlyLeuGlyAsnGly----- 869
Qy      361 TCCGGGATCATGTTGATTTCTCAGAGCTTTTGTGCATTTGCGAGTGTCTGTGTTGCC 420
Db      870 -----AsnGlyArgCysThrIleAspLeu---AsnGlyTyrHis---CysValCysG 885
Qy      421 AACATGCTGTGATCTAACTTTGTGATGTGCACAGCTAACATGTATACACGGCATGGGTGG 480
Db      885 In----- 889
Qy      481 ATGGTGCAGACTGTTCACAGACGTACACATTTGGTGGCTGTCTGTTCGTGGGCTGGATC 540
Db      889 xgGlyThrGlyCysAspThrIleSerMetGlyThrGlyCysGlyAsp----- 903
Qy      541 GCTGAGAGGCTCACTAAATTTGGGGGTGTGATGATGTGATGCGCCGGCGGCTGTGCA 600
Db      904 -----GlyLysAspAsn-----AspGlyAspGly- 911
Qy      601 CCAGAGAAACCAACTACAAAGCCGTTTCTATCATGCTCCAGCCACAGATGTGGCTAC 660
Db      912 -----LeuValAspCysMetAspProAspCysGlyLeuG 923
Qy      661 AA 662
Db      923 In 923

RESULT 34
A36324
growth arrest-specific protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 25-Jan-1991 #sequence_revision 25-Jan-1991 #text_change 09-Jul-2004
C:Accession: A36324
R:Maniatis, G.; Ruato, M.E.; Del Sal, G.; Philipson, L.; Schneider, C.
Mol. Cell. Biol. 10, 2924-2930, 1990
A:Title: A growth arrest-specific (gas) gene codes for a membrane protein.
A:Reference number: A36324; WUID:90258882; PMID:1692961
A:Accession: A36324
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-144 <MAN>
A:Cross-references: UNIPROT:P16646; UNIPARC:UPI00001456E3; GB:M32240
C:Superfamily: growth arrest-specific protein
C:Keywords: transmembrane protein

Alignment Scores:
Pred. No.: 6.4 Length: 144
Score: 96.50 Matches: 44
Percent Similarity: 39.2% Conservative: 21
Best Local Similarity: 26.5% Mismatches: 51
Query Match: 6.2% Indels: 50
DB: 2 Gaps: 8

US-10-537-002-7 (1-786) x A36324 (1-144)
Qy      61 ATCATGTCGACCTGCATGACGACAGTGG-----AGCACCCAGACTTG 105
Db      19 LeupheValSerThrIleValSerGlnThrLeuValGlyAsnGlyIleThrThrAspLeu 38
Qy      106 TACAAACACCCCGTAACAGCTGTTTCAACTACAGAGGCTGTGGCTTCCTGTTCGA 165
::: ||| |||::: |||
```

```
Db      39 TrpIleAsnCysThrThrSerAlaLeu-----GlyAlaValGlnHisCysTyr--- 54
Qy      166 GAGAGCTGTGGCTTACCCAGAGTGCAGGCGGAGTAACTTACCCCTGTGGGCTGCAGCATG 225
Db      55 ---SerSerSerValSerGlu-----Tyr 61
Qy      226 CTGCAGGAGTGCAGAGCCCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 285
Db      62 LeuGlnSerValGlnAlaThrMetIleLeuSerValIlePheSerValLeuAlaLeuPhe 81
Qy      286 GTATCATCTTTGGCTTGAATATGATCCGCATTTGGCAGACATGAGAGATCTGCCAAACC 345
Db      82 LeuphePheCysGlnLeu----- 87
Qy      346 AACATGACATGCAGCTCCGGGATCATGTTGATGCTCTAGGCTTTGTGCATTT---GCT 402
Db      88 ---PheThrLeuThrIleGlyGlyValArgPheTyrIleThrGlyPhePheGlnIleLeuAla 106
Qy      403 GGAGTGTCTGTGTTTGCACAACTGCTGATCTAACTTTGTGATGTCCAGCTAACATG 462
Db      107 GlyLeuGlyVal-----MetSerAlaAlaIle 116
Qy      463 TACACCGCATGCGGTGGATGTGCAGACTGTTCAGACAGGTATACATTTGTGCTGGCT 522
Db      117 TyrThr---ValArgHisSerGlnThrIleValLeuThrAspTyrSerTyrIlePheAla 135
Qy      523 CTGTTCGTGGGCTGGG 538
Db      136 -ThrSerThrProGly 140

RESULT 35
A35524
high-affinity branched-chain amino acid transport system permease protein 11vH [imported]
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: A35524
R:DelVecchio, V.G.; Kaprat, V.; Redkar, R.U.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessac Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: A35524; PMID:11756688
A:Accession: A35524
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-292 <KUR>
A:Cross-references: UNIPROT:Q8YDQ4; UNIPARC:UPI000058368; GB:AE008918; PDB:1AAL53362.1;
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME110121
A:Map position: II

Alignment Scores:
Pred. No.: 6.03 Length: 292
Score: 96.50 Matches: 37
Percent Similarity: 43.8% Conservative: 20
Best Local Similarity: 28.5% Mismatches: 38
Query Match: 6.2% Indels: 35
DB: 2 Gaps: 7

US-10-537-002-7 (1-786) x A35524 (1-292)
Qy      193 GGCCTACTTCAACCTCTGGGCTGCAGCCATGCTG-----CAGCAGTGCAGACCCCTG 246
Db      19 GlyAlaTyrAlaIleLeuGlyValCysAlaIlePheThrTyrArgLeuValAlaVal 38
Qy      247 ATGATCGTAGGATGCTCTGGGTGTCATTTGGC-----CTCCGTGATCATCTTTGGC 300
Db      39 AenPheThrGlyAlaIleGlyAlaIleGlyAlaIleGlyThrPheIleMetValAlaLeuPheGly 58
Qy      301 CTGAATGATCCGATTTGGAGATGAGAGACTGTGCCAAGCCAAACATGACATGAC 360
Db      59 -----AlaGlyValProLeuVal 64
```

Db 626 rghsLaGIuThserGylValAlProAlaGlyLeuThrArgTyLeuProIles 646

Qy 383 CAGGCTTTGTCGCAATGTGTGAGTGTCTGTGTTCGCCAACAGTGTGACTTAATCTT 442

Db 646 erglyValSerSerValAlaAlaLeuSerProTyValAsnLysThrIleThrlGlyAspC 666

Qy 443 GGATGTCCAGAGCTAACAGTATGACCGGATGAGTGTGG-----ATGGTCAGA 490

Db 666 yLeuProlIleuAspMetGluThrlGlyAsnIleGlyAlaTyValValleuValaspG 686

Qy 491 CTGTTTCAGACCGAGTACATTTGGTCGGCTGTCTGTTCGGCTGG 537

Db 686 lInThrlGlyAsnMetAlaThrArgLeuArgAlaAlaValProGlyTyr 701

RESULT 37

JC5732

tumor-associated membrane protein XMP - human

C:Species: Homo sapiens (man)

C:Date: 09-Dec-1997 #sequence\_revision 09-Dec-1997 #text\_change 18-Sep-1998

C:Accession: JC5732

R:Ben-Porath, I.; Benvenisty, N.

G:Gene 183, 69-75, 1996

A:Title: Characterization of a tumor-associated gene, a member of a novel family of genes

A:Reference number: JC5730; MUID:97149281; PMID:8996089

A:Accession: JC5732

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-167 <BEN>

A:Cross-references: UNIPARC:UPI0000178CAD; GB:U52100

C:Comment: This protein is involved in tumor formation and is embryogenesis.

C:Genetics:

A:Gene: XMP

C:Superfamily: growth arrest-specific protein

C:Keywords: glycoprotein

F:2-23/Domain: transmembrane #status predicted <TM1>

F:68-84/Domain: transmembrane #status predicted <TM2>

F:95-117/Domain: transmembrane #status predicted <TM3>

F:143-163/Domain: transmembrane #status predicted <TM4>

F:144,47,52/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:

Pred. No.: 6.94 length: 167

Score: 96.00 Matches: 48

Percent Similarity: 38.5% Conservative: 27

Best Local Similarity: 24.6% Mismatches: 74

Query Match: 6.2% Indels: 46

DB: 2 Gaps: 10

US-10-537-002-7 (1-786) x JC5732 (1-167)

Qy 25 TTGGGATTCGATGTTTCA-----CTGATTTGGATTCGGGACATCATGTGTCACCACTGC 78

Db 5 LeuAlaPheIleIleAlaPheIleIleThrSerAlaAlaLeuPheIleAlaThrVal 24

Qy 79 ATGACCACTGAGAGACCCACAGACTTGTACACACACCCGTAACAGCTGTTTCACTAC 138

Db 25 AspAsnAlaTyrTrpValGlyAspGluPhe-----Phe 35

Qy 139 CAGGGGCTGTGGCGCTCTCGTGTGCGAGAGAGCTGTGGCTTACCGAGTGC----- 189

Db 36 AlaAspValTyrPArgIleLeuThrAsnAsn-----ThrAsnGlyThrValIle 51

Qy 190 -----CGGGGCTACTTCACTTCCTCGTGTGGGCTGCCAGCCATGTCGAGCAGTG 237

Db 52 AsnAspSerPhePheGlnGluTyrSerThr-----LeuGlnAlaVal 64

Qy 238 CGAGCCCTGATGATGTGTAAGGACATGCTCTGGGGTGGCATTGAGCTCTGGATTCATCTTT 297

Db 65 GlnAlaThrMetCilleLeuSerIleIleLeuCySylAlaAlaPhePhe-----IlePhe 82

Qy 298 GCCCGGAATGATCCGATTTGGCAGCATGAGGAAGTCCGCAAGGCAACATGACATG 357

Db 83 ValLeuGlnLeuPheArgLeuGlyGlnGlyGluPro-----PheValLeu 97

[illegible]

Db 134 -----CysCybLysProCysCysGlnSerSerCysCybLysProCysCy 150  
|||  
Qy 419 CCACATGCTGTGAATACTTGTGATGTCCACAGCTAACATGATCACCGGCATGGGTG 478  
|||  
Db 150 e6Ln----- 151

Qy 479 GGATGTCAGACTGTCTTGACAGACCAGGATACATTGGTGGCTCTGTTGGTGGCTGG 538  
:::  
Db 152 ----SerserCysCybLysProCysCysSerserGlyCysGlySer----- 165  
:::  
Qy 539 TCGCTGAGAGCCTCACTAATAATTGGGGGTGTGATGTGATCATGCGCTGCCGGGGCTGG 598  
|||  
Db 166 -----Sercys----- 167

Qy 599 CACCAAGAAACAACCTACAAAAGCCGTTTCTTATCATGCGCTCAGGCCACAGTGTGC 656  
|||||  
Db 168 -----CysGlnSer-----SercysCybLysProCysCysCys 178

RESULT 39  
AG3140  
hypothetical protein Atu4747 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
C:Accession: AG3140  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan,  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.  
ster, E.W.  
AltTitle: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MOID:21608550; PMID:11743193  
A:Accession: AG3140  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <KUR>  
A:Cross-references: UNIPROT:Q8U6F0, UNIPARC:UPI00000D26B1; GB:AE006689; PIDN:AAL45541.1;  
A:Cross-referential source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu4747  
A:Map position: linear chromosome  
C:Superfamily: l-arabinose transport system permease arah

Alignment Scores:

| Pred. No.:             | 6.56  | Length:       | 331 |
|------------------------|-------|---------------|-----|
| Score:                 | 96.00 | Matches:      | 46  |
| Percent Similarity:    | 42.8% | Conservative: | 31  |
| Best Local Similarity: | 25.6% | Mismatches:   | 55  |
| Query Match:           | 2     | Gaps:         | 9   |
| DB:                    |       |               |     |

US-10-537-002-7 (1-786) x AG3140 (1-331)

Qy 193 GGCTACTTACCCCTGCTGGGGCTGCCAGCAGCTCTGACGAGCATGCGAGCCCTGATATC 252  
|||||  
Db 165 GThyHsgLyThrLeuLeuGlyLeuProIleProLeu-----LeuIlePheile 180  
|||||  
Qy 253 GAGAGCATGCTCTGGGTGCCATTTGGCTTCCTGTATTCATCTTTGGCTCGAATGATC 312  
|||||  
Db 181 ValCysValGlyIleTrpHisValLeuLeuThrArgThrIlysLeuGlyPheGlyLeuLeu 200  
|||||  
Qy 313 CGCATTTGGCAGCATGAGGAGCTCTGCCAA-----GCCAATGACATGACCTCC 363  
|||||  
Db 201 MetIleGlySerAsnIleGluAlaIargTySerGlyLeuLeuThrArgGlyIleGln 220  
|||||  
Qy 364 GGGATCATTTATGCTCTCAGGCTTT---TGTGCAATTGCTGGAGTGTCTGTGTTGCC 420  
|||||  
Db 221 ValLeuValTyThrLeuSerGlyLeuMetCysAlaValAlaGlyIleIleMetLeuAla 240  
|||||  
Qy 421 AAC-----ATGCTGGAGCACTAATTC-- 441  
|||||  
Db 241 ArgPheAsnSerValArgValGlyHisGlyGluSerTyrryLeuLeuIleThrValLeuAla 260



442 ..-TGATGTCACAGCTAACATGTAACCGGCGATGGGTGGAGATGGTCGACACTGTTCCAG 498  
 Db .....  
 261 AAlaheLeuGIyGIleAsnProPheGIyGIlePheGIyArGValLeuPro----- 277  
 Oy ACCAGGTACACATTTGGTGGGCTGTGTTCTGTGGGCTGGGTGGAGAGGCTCACACTA 558  
 Db .....-ValPheValAlaLeuIleValLeuGIleLeuLeuSerSerGIyLeuAsnLeu 294  
 278 -----ValPheValAlaLeuIleValLeuGIleLeuLeuSerSerGIyLeuAsnLeu 294  
 Oy ATTGCGG-----GGTGTGATGATGTGCATC----- 582  
 Db .....  
 295 LeuGIyAlaLeuGIleValLeuAlaThrAlaLeuTrpGIyValLeuMetIleValValMet 314  
 Oy GCGTCGCGGGGCGCTGGCAGCCAGAAACCAACTACAAAGCGCTTCTTATCATGCGCTCA 642  
 Db .....  
 315 AlAlaArGIyLeuPheSer-----SerTyPheAlaSer 326  
 RESULT 40  
 E98147  
 hypothetical protein AGR\_L\_265 [imported] - Agrobacterium tumefaciens (strain C58, Cerec  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
 C/Accession: E98147  
 R/Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman,  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 Science 294, 2323-2338, 2001  
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A/Reference number: A97359; MUID:21608551; PMID:11743194  
 A/Accession: E98147  
 A/Status: Preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-331 <KIR>  
 A/Cross-references: UNIPROT:Q8U8R0, UNIPARC:UPI000000D26B1, GB:AE007870; PIDN:AAK8703.1  
 C/Genetics:  
 A/Gene: AGR\_L\_265  
 A/Map position: linear chromosome  
 C/Superfamily: 1-Arabinose transport system permease arah  
 Alignment Scores:  
 Pred. NO.: 6.56 Length: 331  
 Score: 96.00 Matches: 46  
 Percent Similarity: 42.8% Conservative: 31  
 Best Local Similarity: 25.6% Mismatches: 55  
 Query Match: 6.2% Indels: 48  
 Gaps: 9  
 US-10-537-002-7 (1-786) x E98147 (1-331)  
 Oy GGGTACTTCAACCTGCTGGGGCTCCAGCCATGCTGCAGAGTGGAGCCCTGATGATC 252  
 Db .....  
 165 GLyHleGIyHrHleuLeuGIyLeuProIleProIle-----LeuIlePheIle 180  
 Oy GTAGGATCGTCCGGGGTCCATTTGGGCTCTGTGATTCATCTTGGCTGAAATGCATC 312  
 Db .....  
 181 ValCyValGIyIleTrpHleValLeuLeuThrArGIyHrGIyLeuGIyPheGIyLeuLeu 200  
 Oy CGGATGGAGACATGAGGAACTGGCCAA-----GCCAATGACATGACCTCC 363  
 Db .....  
 201 MetIleGIySerAsnIleGluAlaAlaArGIySerGIyLeuAsnTrpArGIySIlleGI 220  
 Oy GGGATCATGTTCAATGTTCTCAGTCTT---TGTCGAATGCTGGAGTGTCTGTGTTGCC 420  
 Db .....  
 221 ValLeuValTyTrHrHleuSerGIyLeuMetCyAlaValAlaGIyIleIleMetLeuAla 240  
 Oy AAC-----ATGCTGGTGACTAACTTC--- 441  
 Db .....  
 241 ArGPheAsnSerValArGIyValGIyHleGIyAsuSerTyLeuLeuIleThrValLeuAla 260  
 Oy ---TGATGTCCACAGTACATGATACACCGGACATGGGTGGAGTGGAGACTGTTCCAG 498  
 Db .....  
 261 AlAlaheLeuGIyGIleAsnProPheGIyGIlePheGIyArGValLeuPro----- 277  
 Oy ACCAGGTACACATTTGGTGGGCTGTGTTCTGTGGGCTGGGTGGAGAGGCTCACACTA 558  
 Db .....-ValPheValAlaLeuIleValLeuGIleLeuLeuSerSerGIyLeuAsnLeu 294  
 278 -----ValPheValAlaLeuIleValLeuGIleLeuLeuSerSerGIyLeuAsnLeu 294  
 Oy ATTGCGG-----GGTGTGATGATGTGCATC----- 582  
 Db .....  
 295 LeuGIyAlaLeuGIleValLeuAlaThrAlaLeuTrpGIyValLeuMetIleValValMet 314  
 Oy GCGTCGCGGGGCGCTGGCAGCCAGAAACCAACTACAAAGCGCTTCTTATCATGCGCTCA 642  
 Db .....  
 315 AlAlaArGIyLeuPheSer-----SerTyPheAlaSer 326

```

Db      278 -----ValPheValAlaLeuIleValLeuGlnLeuSerSeriIleuAsnLeu 294
Oy      559 ATTGGG-----GGTGTGATGATGTGCATC-----582
           ::|||
Db      295 LeuGlyAlaAsnGlnIleLeuAlaThrAlaLeuTrpGlyValLeuMetIleValMet 314
Oy      563 GCCTCCCGGGGCTCTGGACACCAAGAAACCACTACAAACCGCTTCTTATCATCATGCTCA 642
           ::|||
Db      315 AlaAlaArgGlyLeuPheSer-----SeriTrpPheAlaSer 326
           ::|||

RESULT 41
T45918
hypothetical protein F5K20.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #ext_change 05-Oct-2004
C:Accession: T45918
R:Monfort, A.; Casasuberta, B.; Puigdomenech, P.; Mewes, H.W.; Lemcke, K.; May
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223017
A:Accession: T45918
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-361 <MON>
A:Cross-references: UNIPROT:Q9M348; UNIPARC:UPI000095E7C3; EMBL:ALJ32960
A:Experimental source: cultivar Columbia; BAC clone F5K20
C:Genetics:
A:Map position: 3
A:Introns: 93/2; 143/1; 197/3; 279/2; 321/3
A:Note: F5K20.80
C:Superfamily: AAA protein

Alignment Scores:
Pred. No.:          6.51          Length:          361
Score:              96.00          Matches:         59
Percent Similarity: 36.98          Conservative:    45
Best Local Similarity: 20.98       Mismatches:      91
Query Match:        6.28          Indels:          87
DB:                  2            Gaps:            15

US-10-537-002-7 (1-786) x T45918 (1-361)

Oy      21  GGGCTTGGGGTTCGTCGTTTCACTGATTTGGGATTCGGGCATCATTTGTCGCCATGCAT 80
           |||
Db      27  GlyArgGlyValValGlnPheArgSerTrp-----36
Oy      81  GGACCAAGTGAAGACCCCAAGACTTATACAAACCCGTAACAGCTGTTTCACTACCA 140
           |||
Db      37  -----PheProTrpLeuIle-----ProCysPheValValAla 47
Oy      141 GGGGCTGTGGCGCTCTG-----TGTCCGAGAGAGCTTGGC--177
           |||
Db      48  AsnValAlaValPheValIleThrMetTrpValAsnAsnCysProValLys-SerGlyAs 67
Oy      178  ---TTTCAACCGAGTCCGGGGCTACTTCAAC-----CTGCT 209
           |||
Db      67  pCysPheAlaAspPheLeuGlyArgPheSerPheGlnAsnThrArgGlyAsnProLeuLe 87
Oy      210 GGGGCTGCACGACG---CTGCAAGACAGTGCAGGCGCTGATGATGTGAAGCATGCTCT 266
           |||
Db      87  uGlyProSerSerLeuThrLeuGlnThrMetGlyLeuAspValLysValValValLys 107
Oy      267  G-----GGTGCATTGAGCT 281
           |||
Db      107 sGlyAspGlnGlyTrpArgLeuLeuSerCysAsnTrpLeuHsGlyValValHisLe 127
Oy      282 CTTGGATTCATCTTTGGCCCTGAAATGATCCGCAATGGCAGACATGGAAGACTCTGCCA 341
           |||
Db      127 uLeuMetAsnMetLeuThrLeuLeuPheIleGlyLe-----139
Oy      342 AGCCAACATGACACATGACCTCCGGATCATGTTCAATGCTCTTGTGCAATTGC 401
           |||
Db      140  ---ArgMetGlnuLeuArgGlnGlyLeuLeuTrpLeuIleSerGly-----PheG 155

```

er, K.F.X









```

QY 310 ATCCGATTTGGAGCATGAGAGACTCTGCCAAGCCACATGACACTGCCTCCGGGATC 369
Db 163 AsnlyslleGlySerMet-----AsnPhenylrleuMetAsnAsnPh 176
QY 370 ATGTTCAATTGTCTCAGTCTT-----TGTGCATTGTGAGTGCTGTGTTGGCAAC 423
Db 177 MetPheAsnTrpAspLeuLeuTrpPheCysleuLeuCysAlaPheLeuValIysMetPro 196
QY 424 ATGCTGTGACTAATTCTTGATGTCACAGCTAACATGATACCGGACATGGGTGG--- 480
Db 197 MetPheLeuValHisleuTrpLeuProlysaHisValGluAlaProValSerGlySer 216
QY 480 ----- 480
Db 217 MetIleLeuAlaGlyIleMetLeuIysLeuGlyIleTrpGlyMetLeuArgValIleSer 236
QY 481 ATGTGTACAGACTGTTCAGACACAGTACACATTTGTGGCGGCTGTTCGTGGGCTGGATC 540
Db 237 PheLeuGlnLeuMetAsnLeuIysTrpSerPhe-----ValTrpIle 250
QY 541 GCTGAGAGCCTTACACTAATTGGGGGTGTGATGATGTGCATGCGCTGCCGGGCTGGCA 600
Db 251 Ser-----IleSerLeuValGlyIleValLeuValSerLeuValCysIleu----- 265
QY 601 CCAGAAAGAAACCACTACAAAGCC--GTTTCTTATCATGCTCAGGCCAC 648
Db 266 ---ArgGlnThrAspLeuIysAlaLeuIleAlaTrpSerSerValAlaHis 281

RESULT 49
125797
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - fruit fly (Drosophila yakuba) mi
N.Alternate names: NADH-ubiquinone oxidoreductase chain 4
C.Species: mitochondrion Drosophila yakuba
C.Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 09-Jul-2004
C.Accession: J25797
R.Clary, D.O.; Wolstenholme, D.R.
J.Mol. Evol. 22, 252-271, 1985
A.Title: The mitochondrial DNA molecule of Drosophila yakuba: nucleotide sequence, gene
A.Reference number: A92962; MUID:86089137; PMID:3001325
A.Accession: J25797
A.Molecule type: DNA
A.Residues: 1-446 <CLAS>
A.Cross-references: UNIPROT:P07707; UNIPARC:UPI000013075D; GB:X03240; GB:J01400; GB:J014
C.Genetics:
A.Gene: FlyBase:Dyak/mc:ND4
A.Cross-references: FlyBase:FBgn0013185
A.Genome: mitochondrion
A.Genetic code: SGC4
C.Superfamily: NADH dehydrogenase (ubiquinone) chain 4
C.Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Alignment Scores:
Pred. No.: 7.03 Length: 446
Score: 95.50 Matches: 38
Percent Similarity: 41.2% Conservative: 35
Best Local Similarity: 21.5% Mismatches: 57
Query Match: 6.1% Indels: 47
DB: 2 Gaps: 9

US-10-537-002-7 (1-786) x J25797 (1-446)
QY 199 TTCACCTCTGTGGG-----CTGCCAGCATCTGCAGACAGTGGACCGCTGANG 249
Db 125 PheleuIleleuGlyTrpGlyTrpGlnTrpGlnAlaGlnleuGlnAlaGlyValTrpIleuLeu 144
QY 250 ATCGTAGGCATCTCTCGGTGCATTTGGCTCTGTGATCCATCTTTGCGCTGAATGC 309
Db 145 PheTrpTrpIleuLeuVal---SerLeuProMetLeuIleGlyIlePheTrpVal----- 161
QY 310 ATCCGATTTGGAGCATGAGAGACTCTGCCAAGCCACATGACACTGCCTCCGGGATC 369
Db 162 -----MetAsnlyslThrGlySerMetAsnPhenylrleuMetAsnAsnPh 176

```

```

QY 370 ATGTTCAATTGTCTCAGTCTT-----TGTGCATTGTGAGTGCTGTGTTGGCAAC 423
Db 177 MetPheAsnTrpAspLeuLeuTrpPheCysleuLeuCysAlaPheLeuValIysMetPro 196
QY 424 ATGCTGTGACTAATTCTTGATGTCACAGCTAACATGATACCGGACATGGGTGG--- 480
Db 197 MetPheLeuValHisleuTrpLeuProlysaHisValGluAlaProValSerGlySer 216
QY 480 ----- 480
Db 217 MetIleLeuAlaGlyIleMetLeuIysLeuGlyIleTrpGlyMetLeuArgValIleAsn 236
QY 481 ATGTGTACAGACTGTTCAGACACAGTACACATTTGTGGCGGCTGTTCGTGGGCTGGATC 540
Db 237 PheLeuGlnLeuMetAsnLeuIysTrpSerPhe-----ValTrpIle 250
QY 541 GCTGAGAGCCTTACACTAATTGGGGGTGTGATGATGTGCATGCGCTGCCGGGCTGGCA 600
Db 251 Ser-----IleSerLeuValGlyIleValLeuValSerLeuValCysIleu----- 265
QY 601 CCAGAAAGAAACCACTACAAAGCC--GTTTCTTATCATGCTCAGGCCAC 648
Db 266 ---ArgGlnThrAspLeuIysAlaLeuIleAlaTrpSerSerValAlaHis 281

RESULT 50
A25704
synapsin I - rat
C.Species: Rattus norvegicus (Norway rat)
C.Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 09-Jul-2004
C.Accession: A25704
R.McCaferly, C.A.; DeGennaro, L.J.
EMBO J. 5, 3167-3173, 1986
A.Title: Determination and analysis of the primary structure of the nerve terminal speci
A.Reference number: A25704; MUID:87133474; PMID:3028773
A.Cross-references: UNIPROT:P09951; UNIPARC:UPI0000170812; GB:X04655; NID:g57181; PIDN:CF
A.Experimental source: brain
C.Keywords: actin binding; alternative splicing; phosphoprotein

Alignment Scores:
Pred. No.: 6.78 Length: 691
Score: 95.50 Matches: 71
Percent Similarity: 30.7% Conservative: 36
Best Local Similarity: 20.3% Mismatches: 117
Query Match: 6.1% Indels: 125
DB: 2 Gaps: 12

US-10-537-002-7 (1-786) x A25704 (1-691)
QY 729 TTCATCGTATATCTTCTT-----GTTTGTGTGT 700
Db 138 SerIleLeuTrpLeuMetProMetValAspSerProTrpTrpTrpIysPheHeGlyMet 157
QY 699 GGAACCCAAAGCAGTGTGAGCTTCGAGCTTCAGAGCTTGATGAGCAACATGTGGCTGA 640
Db 158 GlySerIysValAlaArgSerLeuIysProAspPheValIleuIleArgGlnHisAlaPhe 177
QY 639 GGCATGATTAAGAAAGCGC-----TTTGTAGTTGGTTTCTTCTGTGCCAGGCC 592
Db 178 SerMetAlaArgAsnGlyAspTrpArgSerLeuValIleGlyLeuGlnIleValIle 197
QY 591 CCGGAGGCGAGTACATCATCATCACACCCCAATTGATGTGAG-----GCC 547
Db 198 ProSerValAsnSerLeuHisSerValIleTrpAsnPheCysAspLysProTrpValAlaPheAla 217
QY 546 TCACAGCAGCCAGCCAGCAAGACAGAGCCGACCAATGTGTATCTGTCTGCAAGATCTG 487
Db 218 GlnMetValArgLeuHisIleIysIleValGlnTrpGlnGlnPheProIleuIleAspGlnThr 237
QY 486 CACCATCCACCCATGCGCGGTGTACATGTAGCTGTGACATCCAGAAATTAGTACAG 427

```

|    |     |   |     |
|----|-----|---|-----|
| Db | 238 | PhenylProteinHisIlysgluMetLeuSerSerThrThyTrpProValValIlyMet     | 257 |
| Oy | 426 | CATGTGGCAACACAGACACTCCAGCAATTCGACAAAGCATCGACCAATGACATGAT        | 367 |
| Db | 258 | -----GlyHisAlaHisSerGlyMetGlyValValIlyValIlySerAlaHisGlnHisAsp  | 274 |
| Oy | 367 | -----   | 367 |
| Db | 275 | PhenAlaPheAlaSerValValAlaLeuThrLysThrTyralAlaThrAlaGluProPhe    | 294 |
| Oy | 366 | -----CCGAGAGTGAAGTGCATGTGGCTTTGGC                               | 337 |
| Db | 295 | IleAspAlaLysTyraAspValArgValGlnLysIleGlyGlnHisGlyLysAlaTyMet    | 314 |
| Oy | 336 | AGAGCTCCACATGCGCCCAT  | 316 |
| Db | 315 | ArgThrSerValSerGlyAsnTrpLysThrAsnThrGlySerAlaMetLeuGlnIle       | 334 |
| Oy | 316 | -----   | 316 |
| Db | 335 | AlaMetSerAspArgTytyLeuTrpValAspThrCysSerGlnIlePheGlyGlyLeu      | 354 |
| Oy | 315 | -----GCGAGTCATTTACGGCGCAAGATGA-----TACGAGAG                     | 280 |
| Db | 355 | AspIleCysAlaValAlaGlnIleAsnIleGlyLysAspGlyArgAspHisIleIleGluVal | 374 |
| Oy | 279 | GCCATGGCACCCAGAGACATGCTCATGCATCATCAGAGGCTCGACATGCTCCGACATGGC    | 220 |
| Db | 375 | ValGlySerSerMetProLeuIleGlyAspHisGlnGlyAspGluAspLysGlnLeuIle    | 394 |
| Oy | 219 | TGGCAG-----CCGACGAGGGGTGAATGAGCCCG                              | 190 |
| Db | 395 | ValGlnLeuValValAsnLysMetThrGlnAlaLeuProArgGlnArgAspAlaSerPro    | 414 |
| Oy | 189 | GCACTCGGTGAAGCCAGACCTCTTCGACACAGAGCGCCACAGCCCTGTATGTGAA         | 130 |
| Db | 415 | GlyArgGly-----HisSerGlnThrProSerProGlyAla-----                  | 427 |
| Oy | 129 | AACAGCTGTACGGGGTTTGTACAAGTCTTGAGTCCACT-----G                    | 85  |
| Db | 428 | -----LeuProLeuGlnTyArgGlnThrSerGlnGlnProAlaG                    | 440 |
| Oy | 84  | GTCATGACAGTGGGACGAATGATGCCCGCCCAATCCCAATCAGTGAACCAACGAAACCCCA   | 25  |
| Db | 440 | LysProProAlaGlnGlnGlnLysProProProGlnGlyLysProProGlnPro--GlyProG | 459 |
| Oy | 24  | GCCTGACAGGACGTCAAGCGCA  | 2   |
| Db | 459 | LysProGlnArgGlnGlnLysProPro                                     | 466 |

Search completed: September 26, 2006, 00:22:03  
Job time : 80.5 secs

**THIS PAGE IS BLANK**



This Page is inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record

## BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images  
problems checked, please do not report the  
problems to the IFW Image Problem Mailbox**

**THIS PAGE IS BLANK**